Display D
UNIVERSITY OF MICHIGAN
NOV 4 1953
MATHEMATICS LIBRARY

AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

REINHOLD BAER
UNIVERSITY OF ILLINOIS

WEI-LIANG CHOW
THE JOHNS HOPKINS UNIVERSITY

SAMUEL EILENBERG
COLUMBIA UNIVERSITY

AUREL WINTNER
THE JOHNS HOPKINS UNIVERSITY

WITH THE COÖPERATION OF

L. V. AHLFORS
S. S. CHERN
C. CHEVALLEY
A. M. GLEASON
P. HARTMAN
G. P. HOCHSCHILD
I. KAPLANSKY
E. R. KOLCHIN
W. S. MASSEY
A. D. WALLACE
ANDRÉ WEIL

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

Volume LXXV, Number 4
OCTOBER, 1953

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

CONTENTS

The AMERICAN JOURNAL OF MATHEMATICS will appear four times yearly.

The subscription price of the JOURNAL is $8.50 in the U. S., Pan American countries, and Spain; $8.75 in Canada; and $9.00 in other foreign countries. The price of single numbers is $2.50.

Papers intended for publication in the JOURNAL may be sent to any of the Editors.

Editorial communications should be sent to Professor AUREL WINTNER at The Johns Hopkins University.

Subscriptions to the JOURNAL and all business communications should be sent to THE JOHNS HOPKINS PRESS, BALTIMORE 18, MARYLAND, U. S. A.

Entered as second-class matter at the Baltimore, Maryland, Postoffice, acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917, Authorized on July 3, 1918.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

ON THE STRUCTURE OF UNITARY GROUPS (II).*

By JEAN DIEUDONNÉ.

1. This paper adds some miscellaneous results on unitary groups to those which have been proved in [4]. In sections 2 to 6, I show how the study of unitary groups over a field of characteristic 2 can always be reduced to the case, considered in [4], in which the hermitian form is "trace-valued." In section 8 to 12 I prove that, with a single exception, quasi-symmetries generate the unitary group, and deduce from that fact certain information on the determinant of a unitary transformation.

2. The terminology and notations are those of [4]. When K is a sfield of characteristic 2 and f an arbitrary nondegenerate hermitian form over the n-dimensional space E, we are going to see that the structure of the group $U_n(K, f)$ can essentially be reduced to that of another unitary group $U_m(K, f_1)$, where f_1 is a "trace-valued" form, that is, such that every value $f_1(y, y)$ in K can be written $\xi + \xi^J$. We observe here that the case in which K is commutative and J the identity is included in what follows, and gives back the treatment of the groups leaving invariant a symmetric form over a field of characteristic 2 ([2], p. 60), of which the following is obviously a generalization.

Let V be the subset of E consisting of vectors x such that $f(x, x)$ has the form $\xi + \xi^J$; owing to the formulas

$$f(x + y, x + y) = f(x, x) + f(y, y) + f(x, y) + (f(x, y))^J$$

and

$$f(x\lambda, x\lambda) = \lambda^J f(x, x)\lambda$$

V is a vector subspace of E. Let V^* be the subspace orthogonal to V, $V_1 = V \cap V^*$, V_2 a subspace of V supplementary to V_1; let q be the dimension of V_1, m that of V_2. V^* has, then, dimension $n - (m + q)$; let V_3 be a subspace of V^* supplementary to V_1, of dimension $n - m - 2q$. V_2 and V_3 are non-isotropic, and are orthogonal to each other, therefore $V_2 + V_3$ is non-isotropic; so is therefore $(V_2 + V_3)^*$, which has dimension $2q$ and is supplementary to $V_2 + V_3$. For future purposes, we prove the following lemma:

* Received January 29, 1953.

LEMMA 1. *There exists a basis* e_i $(1 \leq i \leq 2q)$ *of* $(V_2 + V_3)^* = V_2^* \cap V_3^*$ *such that the vectors* $e_1, e_2, \cdots, e_q$ *form a basis of* V_1 *and that* $f(e_i, e_{q+j}) = 0$ *for* $i \neq j$, $f(e_i, e_{q+i}) = 1$ *for* $1 \leq i \leq q$.

It is clear that V_1 is contained in $V_2^* \cap V_3^*$. Let $e_1 \neq 0$ be an arbitrary vector in V_1; e_1 cannot be orthogonal to $V_2^* \cap V_3^*$, for as it is already orthogonal to $V_2 + V_3$, it would be orthogonal to E, which is contrary to the assumption that f is nondegenerate. There is therefore a vector e_{q+1} in $V_2^* \cap V_3^*$ such that $f(e_1, e_{q+1}) \neq 0$, and as e_1 is orthogonal to V_1, e_{q+1} is not in V_1; by multiplication of e_{q+1} by a scalar, we can suppose that $f(e_1, e_{q+1}) = 1$. As $f(e_1, e_1) = 0$, it is readily verified that the restriction of f to the plane P generated by e_1 and e_{q+1} is nondegenerate. Therefore the $(n-2)$-dimensional subspace P^* orthogonal to P is supplementary to P; moreover, P^* contains $V_2 + V_3$, and the hyperplane H generated by P^* and e_1 is orthogonal to e_1 and contains V_1; it follows from this that the intersection $P^* \cap V_1$ is $(q-1)$-dimensional. We can then prove the lemma by induction on q, since the restriction of f to P^* is nondegenerate, and the subspace of P^* where $f(x, x)$ is "trace-valued" consists of $V \cap P^* = V_2 + (V_1 \cap P^*)$; there exists therefore a basis $e_2, \cdots, e_q, e_{q+2}, \cdots, e_{2q}$ of $P^* \cap (V_2^* \cap V_3^*)$ such that $e_2, \cdots, e_q$ form a basis for $P^* \cap V_1$, and $f(e_i, e_{q+j}) = \delta_{ij}$ for $i \geq 2$ and $j \geq 2$; it is then clear that e_i $(1 \leq i \leq 2q)$ verify the conditions of the lemma.

We shall designate by V_4 the subspace generated by $e_{q+1}, \cdots, e_{2q}$; E is therefore the direct sum of the 4 subspaces V_1, V_2, V_3, V_4.

3. Let u be an arbitrary transformation in the unitary group $U_n(K, f)$; for every $x \in V_3 + V_4$, we can write $u(x) = v(x) + w(x)$, where $v(x) \in V_3 + V_4$ and $w(x) \in V = V_1 + V_2$; the relation $f(u(x), u(x)) = f(x, x)$ can be written

$$f(v(x), v(x)) + f(w(x), w(x)) + f(v(x), w(x)) + f(v(x), w(x))^J = f(x, x)$$

hence

$$f(v(x) - x, v(x) - x) = f(v(x), w(x) - x) + (f(v(x), w(x) - x))^J + f(w(x), w(x)).$$

But as $w(x) \in V$, $f(w(x), w(x))$ can be written $\lambda + \lambda^J$ by assumption, hence the same conclusion is true of $f(v(x) - x, v(x) - x)$, from which it follows, by definition, that $v(x) - x \in V$; but $v(x) - x$ is by assumption in $V_3 + V_4$, which proves that $v(x) = x$ in $V_3 + V_4$, hence $u(x) = x + w(x)$.

We next remark that u leaves V invariant (globally) by definition,

hence it also leaves V^* globally invariant, and the same is true therefore of $V_1 = V \cap V^*$. Let $x \in V_1$, $y \in V_4$; the relation $f(u(x), u(y)) = f(x, y)$ yields

$$f(u(x), y + w(y)) = f(x, y)$$

and as $w(y) \in V$ is orthogonal to $u(x) \in V_1$, this is equivalent to $f(u(x) - x, y) = 0$. In other words, $u(x) - x$ is a vector in V_1 which is orthogonal to V_4; but such a vector is therefore orthogonal to E (since V_1 is orthogonal to $V_1 + V_2 + V_3$), hence 0; in other words, $u(x) = x$ for every $x \in V_1$.

Finally, for $x \in V_3$, we must have $u(x) = x + w(x) \in V^*$, hence $w(x) \in V^*$; but as $w(x) \in V$, we have $w(x) \in V_1$; for every $y \in V_4$, we write $f(u(x), u(y)) = f(x, y)$, or equivalently

$$f(x + w(x), y + w(y)) = f(x, y).$$

As x is orthogonal to $w(y) \in V$, and $w(x) \in V_1$ is also orthogonal to $w(y)$, this relation reduces to $f(w(x), y) = 0$ for every $y \in V_4$; the same argument as above proves then that $w(x) = 0$.

Summing up, we see that $u(x) = x$ in $V^* = V_1 + V_3$, and that in V_4 we may write $u(x) = x + w_1(x) + w_2(x)$, with $w_1(x) \in V_1$ and $w_2(x) \in V_2$.

4. As u leaves invariant (globally) $V = V_1 + V_2$, for $x \in V_2$ we can write $u(x) = u_0(x) + v_1(x)$, where $u_0(x) \in V_2$ and $v_1(x) \in V_1$; if $y \in V_2$, the relation $f(u(x), u(y)) = f(x, y)$ yields (owing to the fact that V_1 is orthogonal to V)

$$f(u_0(x), u_0(y)) = f(x, y).$$

This shows that u_0 is a mapping of V_2 into itself which belongs to the unitary group $U_m(K, f_1)$, where f_1 is the restriction of f to V_2; due to the definition of V_2, f_1 is *nondegenerate* and *trace-valued*, and therefore the study of the group $U_m(K, f_1)$ can be made with the methods developed in [4].

In order to write that u satisfies the identity $f(u(x), u(y)) = f(x, y)$ for all pairs of vectors x, y in E, it remains only to write this identity when $x \in V_2$ and $y \in V_4$, or when both x and y are in V_4; this yields the two relations

(1) $f(v_1(x), y) + f(u_0(x), w_2(y)) = 0$ for $x \in V_2$ and $y \in V_4$;

(2) $f(x, w_1(y)) + f(w_1(x), y) + f(w_2(x), w_2(y)) = 0$ for $x \in V_4$, $y \in V_4$.

We want to prove the following

LEMMA 2. *For an arbitrary unitary mapping $u_0 \in U_m(K, f_1)$, and an arbitrary linear mapping v_1 from V_2 into V_1, there exists at least one unitary mapping $u \in U_n(K, f)$ whose restriction to V_2 is $u_0 + v_1$.*

Let us prove first that equation (1) determines entirely the linear mapping w_2 of V_4 into V_2. Indeed, for any given $y \in V_4$, the mapping $x \to f(y, v_1(x))$ is a linear form over the vector space V_2; as the restriction of f to V_2 is a nondegenerate bilinear hermitian form, there is one and only one vector $z \in V_2$ such that $f(y, v_1(x)) = f(z, x)$ for all $x \in V_2$, and it is clear that z is a linear function of $y \in V_4$; we can therefore write identically $f(y, v_1(x)) = f(h(y), x)$, where h is a linear mapping of V_4 into V_2. On the other hand, we have $f(h(y), x) = f(u_0(h(y)), u_0(x))$ since $h(y) \in V_2$; this shows that $w_2(y) = u_0(h(y))$, and proves our assertion.

We have now to determine w_1 from equation (2); it will be enough to verify the equations

$$f(e_{q+i}, w_1(e_{q+j})) + f(w_1(e_{q+i}), e_{q+j}) = f(w_2(e_{q+i}), w_2(e_{q+j})) \tag{3}$$

for all values of $i \leq q$ and $j \leq q$. Let $w_1(e_{q+i}) = \sum_{j=1}^{q} e_j \alpha_{ij}$; equations (3) become then

$$\alpha_{ji} + \alpha_{ij}{}^J = \rho_{ij} \tag{4}$$

where $\rho_{ij} = f(w_2(e_{q+i}), w_2(e_{q+j}))$. Now observe that as w_2 maps V_4 into V_2, ρ_{ii} can be written as a " trace " $\gamma_i + \gamma_i{}^J$; on the other hand $\rho_{ji} = \rho_{ij}{}^J$. We can therefore solve equations (4) by taking for instance $\alpha_{ii} = \gamma_i$ and $\alpha_{ij} = \rho_{ij}{}^J$ for $i < j$, $\alpha_{ij} = 0$ for $i > j$. Lemma 2 is thus completely proved.

5. For every $u \in U_n(K, f)$, let u_V be the restriction of u to the subspace V; the mapping $u \to u_V$ is a homomorphism of U_n onto a group G_V of linear mappings of V into itself, leaving every element of V_1 invariant. We verify immediately that in G_V the subgroup G_0 consisting of the u_V such that $u_0(x) = x$ in V_2 is an abelian normal subgroup of G_V, isomorphic to the additive group of all linear mappings of V_2 into V_1, or equivalently to the additive abelian group K^{mq}. G_0 is the kernel of the homomorphism $u_V \to u_0$, and this homomorphism maps G_V on the group $U_m(K, f_1)$, since it follows from Lemma 2 that u_0 may be chosen arbitrarily in that last group; therefore G_V/G_0 is isomorphic to $U_m(K, f_1)$. Finally, the kernel of $u \to u_V$ is the normal subgroup Γ of U_n consisting of the transformations u such that $v_1(x) = 0$ and $u_0(x) = x$ for $x \in V_2$; equation (1) then yields $f(x, w_2(y)) = 0$ for $x \in V_2$ and $y \in V_4$; as $w_2(y) \in V_2$ and the restriction of f to V_2 is nondegenerate, $w_2(y) = 0$ for all $y \in V_4$. Equation (2) then reduces to

$$f(x, w_1(y)) + f(w_1(x), y) = 0$$

for $x \in V_4$ and $y \in V_4$; and the same argument as in the proof of Lemma 2 shows that this equation is equivalent to the relations

$$\alpha_{ji} + \alpha_{ij}{}^J = 0 \tag{5}$$

for the elements of the matrix of w_1. This determines the elements α_{ji} for $i < j$ when the elements α_{ij} for $i < j$ are taken arbitrarily in K; on the other hand, relation (5) for $i = j$ shows that α_{ii} must belong to the set S of symmetric elements of K. As it is readily seen that the group Γ is isomorphic to the additive group of the matrices of the mappings w_1, this group is therefore isomorphic to $S^q \times K^{q(q-1)/2}$.

Summing up the preceding results, we finally have

THEOREM 1. *The group $U_n(K, f)$ has a composition series $U_n \supset \Gamma_0 \supset \Gamma$ such that $U_n(K, f)/\Gamma_0$ is isomorphic to $U_m(K, f_1)$, where $m \leq n$ and f_1 is a nondegenerate, trace-valued hermitian form; Γ_0/Γ is an abelian group isomorphic to the additive group K^{mq} and Γ is an abelian group isomorphic to the additive group $S^q \times K^{q(q-1)/2}$ $(2q \leq n - m)$.*

6. In addition to the condition $m + 2q \leq n$, the numbers m and q may have to satisfy additional restrictions due to the nature of the sfield K. For instance, suppose K is the reflexive sfield of rank 4 over the field $Z = \mathfrak{F}_2(X)$ (field of rational functions of one indeterminate over the prime field $\mathfrak{F}_2$ of 2 elements), with basis 1, ω, θ and $\theta\omega$ such that $\omega^2 = \omega + 1$, $\theta^2 = X$ and $\theta\omega\theta^{-1} = \omega + 1$, ([2], p. 73); 1, θ and $\theta\omega$ constitute a basis for the set of symmetric elements, and the "traces" $\xi + \xi^J$ are the elements of Z. Now, as $f(x, x)$ must not be equal to a trace in $V_3 + V_4$, except for $x = 0$, and in particular $f(x, x) \neq 0$ for $x \neq 0$ in $V_3 + V_4$, a classical argument shows that there exists in $V_3 + V_4$ an orthogonal basis (c_i) $(1 \leq i \leq n - m - q)$, that is, such that $f(c_i, c_j) = 0$ for $i \neq j$. In order that $f(x, x)$ be distinct from a trace for $x \neq 0$ in $V_3 + V_4$, it is then necessary and sufficient that $f(c_i, c_i) = a_i\theta + b_i\theta\omega$, with a_i and b_i in Z and not both 0, and that the equations $\sum_{i=1}^{n-m-q} a_i N(\xi_i) = 0$ and $\sum_{i=1}^{n-m-q} b_i N(\xi_i) = 0$ have no solution other than $\xi_i = 0$ in the sfield K. But Z is a quadratic inseparable extension of its subfield $Z^2 = \mathfrak{F}_2(X^2)$, so that we may write $a_i = p_i^2 + Xq_i^2$, $b_i = r_i^2 + Xs_i^2$, with p_i, q_i, r_i, s_i in Z. Now if we take ξ_i in Z, $N(\xi_i) = \xi_i^2$, and the equations $\sum_i a_i N(\xi_i) = 0$ and $\sum_i b_i N(\xi_i) = 0$ are equivalent to the system of 4 linear equations in the ξ_i, $\sum_i p_i\xi_i = 0$, $\sum_i q_i\xi_i = 0$, $\sum_i r_i\xi_i = 0$, $\sum_i s_i\xi_i = 0$. Such a system has a nontrivial solution in Z as soon as $n - m - q > 4$, hence we

get the inequality $m + q \geq n - 4$, which, coupled with $m + 2q \leq n$, yields in particular $q \leq 4$ and $m \geq n - 8$.

7. Theorem 1 makes it possible to study the centralizer of an *involution* s in a group $U_n(K, f)$, where K has characteristic 2 and f is a nondegenerate trace-valued hermitian form. Such a transformation has the form $x \to x + t(x)$, where $x \to t(x)$ is a linear mapping of E onto a totally isotropic subspace U of dimension $p \leq n/2$, whose orthogonal subspace $U^* \supset U$ is equal to the kernel $t^{-1}(0)$ (invariant elements of s). Let V be a non-isotropic $(n - 2p)$-dimensional subspace, supplementary to U in U^* and let W be a totally isotropic p-dimensional subspace supplementary to U in the subspace V^* orthogonal to V.

The determination of the centralizer Γ of s follows the same method as in ([3], section 25, pp. 35-37), and we refer therefore the reader to that paper for detailed proofs. We first consider the normal subgroup Γ_1 of Γ consisting of the transformations u leaving invariant every element in U, and the normal subgroup $\Gamma_2 \subset \Gamma_1$ of Γ, consisting of transformations leaving invariant every element of U^*; it is easily verified that Γ_2 is an abelian group. The factor group Γ_1/Γ_2 is isomorphic to the group Γ_1' consisting of the restrictions to U^* of transformations belonging to Γ_1. Within this last group we consider the normal subgroup Γ_2' consisting of the elements which leave invariant every class mod. U in the space U^*, and Γ_2' is again abelian.

On the other hand, for any element $v \in \Gamma_1'$ and every $z \in V$ we may write $v(z) = v'(z) + v''(z)$, where $v'(z) \in V$ and $v''(z) \in U$; $v \to v'$ is a homomorphism of Γ_1' onto a group Γ'' of transformations of V, which is seen to be identical with $U_{n-2p}(K, f_1)$, where f_1 is the restriction of f to V; moreover, the kernel of $v \to v'$ is Γ_2', so that Γ_1'/Γ_2' is isomorphic to $U_{n-2p}(K, f_1)$.

Finally, for every $u \in \Gamma$ and every $x \in W$, let $w(x)$ be the component of $u(x)$ on W (in the decomposition of E as a direct sum of U^* and W); let $\bar{u}(x) = w(x)$ for $x \in W$, $\bar{u}(x) = u(x)$ for $x \in U$ and $\bar{u}(x) = x$ for $x \in V$. It can be verified that $\bar{u} \in \Gamma$, is in the same class as u mod. Γ_1, and that $u \to \bar{u}$ is a homomorphism of Γ onto a subgroup $\bar{\Gamma}$ of Γ, isomorphic to Γ/Γ_1, consisting of transformations leaving invariant every element of V (and which can therefore be considered as unitary transformations in the nonisotropic subspace V^* of E), and leaving invariant (globally) U and W. If we take a basis of V^* consisting of p vectors e_i in U and p vectors e_{p+i} in W, such that $f(e_i, e_{p+j}) = \delta_{ij}$ (see [2], p. 6 and [4], p. 369), the matrix of $\bar{u}$ has then the form $\begin{pmatrix} A & 0 \\ 0 & {}^t\!A^{-1} \end{pmatrix}$ where A is a square matrix of order p and ${}^t\!A$ its

contragredient. For that same basis, the matrix of s is $\begin{pmatrix} I & S \\ 0 & I \end{pmatrix}$, where S is a hermitian matrix of order p (i. e., such that ${}^tS = S^J$) which is non-degenerate, but can correspond to a non-trace-valued form f_2. The condition of permutability for $\bar{u}$ and s then reduces to $AS({}^tA^J) = S$, and this proves of course that the group $\bar{\Gamma}$ is isomorphic to $U_p(K, f_2)$. Theorem 1 now shows that this group has a composition series in which the factor groups are abelian, with the exception of one group isomorphic to $U_m(K, f_3)$, where $m \leq p$ and f_3 is a trace-valued form.

Summing up, we conclude that *the group* Γ *has a composition series in which all factor groups are abelian, with the exception of two groups isomorphic to* $U_{n-2p}(K, f_1)$ *and* $U_m(K, f_3)$, *where* $m \leq p$ (possibly $m = 0$, in which case $U_m(K, f_3)$ is understood to be reduced to the identity), *and where* f_1 *and* f_3 *are trace-valued forms.*

8. In this section, K is an arbitrary sfield with an involution J, and f a *hermitian* nondegenerate form over an n-dimensional right vector space E over K; f is supposed to be trace-valued when K has characteristic 2, and we exclude the case of the symplectic groups over fields of characteristic 2; in other words, we always suppose there are vectors $x \in E$ such that $f(x, x) \neq 0$. For such a nonisotropic vector a, a *quasi-symmetry* of hyperplane H orthogonal to a is a unitary transformation of $U_n(K, f)$ leaving invariant every element of H; there are always quasi-symmetries of hyperplane H not reduced to the identity ([4], p. 370).

THEOREM 2. *The group* $U_n(K, f)$ *is generated by quasi-symmetries except when* $n = 2$, $K = \mathfrak{F}_4$ *and* $J \neq 1$.

Of course, for orthogonal groups over fields of characteristic $\neq 2$, this reduces to the well-known theorem that the group is generated by symmetries, ([2], p. 20, prop. 8).

The theorem being obvious for $n = 1$, we use induction on n. Let $u \in U_n(K, f)$, and suppose we can find a nonisotropic vector x and a scalar $\alpha \in K$ with the following properties: 1° $f(x\alpha, x\alpha) = \alpha^J f(x, x)\alpha = f(x, x)$; 2° $u(x) - x\alpha$ is not isotropic. Then, if H is the hyperplane orthogonal to $u(x) - x\alpha$, one can write $x\alpha = y + z$, where $y \in H$ and z is orthogonal to H, and similarly $u(x) = y + z'$, where z' is orthogonal to H; as $f(u(x), u(x)) = f(x, x) = f(x\alpha, x\alpha)$, we have also $f(z, z) = f(z', z')$, hence, as $z' = z\gamma$, $\gamma^J f(z, z)\gamma = f(z, z)$. It follows that the quasi-symmetry s of hyperplane H, transforming z into $z\gamma = z'$, will transform $x\alpha$ into $u(x)$. A similar argument

proves that there is a quasi-symmetry s', whose hyperplane is orthogonal to x, and which transforms x into $x\alpha$; then $(ss')^{-1}u$ will leave x invariant, and can be considered as a unitary transformation operating in the hyperplane orthogonal to x. As such, it is a product of quasi-symmetries by the induction hypothesis, and so therefore is u.

If the characteristic of K is not 2, then for every nonisotropic vector x, one of the scalars 1 or -1 will satisfy both conditions 1° and 2°, for $u(x) - x$ and $u(x) + x$ cannot both be isotropic, as this would imply $f(x,x) + f(u(x), u(x)) = 0$, or $2f(x,x) = 0$, contrary to assumption. We can therefore restrict our further arguments to the case in which K *has characteristic* 2 and $J \neq 1$ (otherwise, f, being trace-valued, would be an alternate form), and for every *nonisotropic vector* x in E, $u(x) - x$ is *isotropic*; this condition is obviously equivalent to the relation

$$f(x, u(x)) + f(u(x), x) = 0$$

or $f(x, u(x)) = (f(x, u(x)))^J$. We are going to show that in such a case, the last equation is also true for every *isotropic* vector x in E, unless $K = \mathfrak{F}_4$ and $J \neq 1$ (in which case the field K_0 invariant by J is the field $\mathfrak{F}_2$ of two elements).

Let a be an arbitrary isotropic vector in E; there exists a second isotropic vector b such that $f(a,b) = 1$. Let $x = a\zeta + b$; if we write that x is isotropic, we obtain the relation $\zeta + \zeta^J = 0$, in other words, ζ must be a *symmetric* element. By assumption,

$$f(a\zeta + b, u(a)\zeta + u(b)) = \zeta^J f(a, u(a))\zeta + \zeta^J f(a, u(b)) + f(b, u(a))\zeta + f(b, u(b))$$

is a *symmetric* element when ζ is *not symmetric*; if we write $f(a, u(a)) = \alpha$, $f(b, u(b)) = \beta$, $f(a, u(b)) = \lambda$, $f(b, u(a)) = \mu$ we can also say that we have the relation

$$\zeta^J(\alpha + \alpha^J)\zeta + \zeta^J(\lambda + \mu^J) + (\mu + \lambda^J)\zeta + (\beta + \beta^J) = 0 \tag{6}$$

for every *nonsymmetric* ζ. Now, if ζ is not symmetric, $\xi + \zeta$ is also not symmetric for every *symmetric* ξ; replacing ζ by $\xi + \zeta$ in (6) and subtracting from (6), we get

$$\xi(\alpha + \alpha^J)\xi + \zeta^J(\alpha + \alpha^J)\xi + \xi(\alpha + \alpha^J)\zeta + \xi(\lambda + \mu^J) + (\mu + \lambda^J)\xi = 0 \tag{7}$$

for every nonsymmetric ζ and every symmetric ξ. Replacing now ξ by $\xi + \eta$, where ξ and η are both symmetric, and subtracting, from the resulting equation, equation (7) and the analogue with ξ replaced by η, we obtain

$$\xi(\alpha + \alpha^J)\eta + \eta(\alpha + \alpha^J)\xi = 0 \tag{8}$$

for every pair of symmetric elements ξ, η. In particular, for $\eta = 1$, this shows that $\alpha + \alpha^J$ commutes with every symmetric element of K. But as K has characteristic 2, symmetric elements generate K over its center Z, ([4], p. 367, Lemma 1) hence $\alpha + \alpha^J$ must be in Z. If $\alpha + \alpha^J \neq 0$, (8) shows that any two symmetric elements are permutable, hence K is *commutative*. But then (7) is an equation of degree 2 in ξ, where the coefficient of ξ^2 is not 0, hence it has at most two distinct roots in K, and if the field K_0 of symmetric elements of K has more than 2 elements, we thus reach a contradiction, which proves that $\alpha + \alpha^J = 0$; in other words $f(a, u(a))$ is symmetric for every isotropic vector a.

This being proved, we have now the identity

$$f(x + y\xi, u(x) + u(y)\xi) = f(u(x) + u(y)\xi, x + y\xi)$$

for every pair of vectors x, y and every scalar $\xi \in K$. Writing $\lambda = f(x, u(y))$, $\mu = f(y, u(x))$, this shows that $\lambda\xi + \xi^J\mu$ is symmetric for all $\xi \in K$, in other words $(\lambda + \mu^J)\xi = \xi^J(\mu + \lambda^J)$ for all $\xi \in K$; this can only be the case if $\lambda + \mu^J = 0$, otherwise it would give, with $\rho = \lambda + \mu^J$, first $\rho^J = \rho$ (with $\xi = 1$) and then $\xi^J = \rho\xi\rho^{-1}$ for all $\xi \in K$. The involution J would thus be an automorphism of K, which is only possible if K is commutative, and $J = 1$, a case we have excluded.

From these results, we conclude therefore that the exceptional case occurs only when

$$f(x, u(y)) = (f(y, u(x)))^J \tag{9}$$

or equivalently

$$f(x, u(y)) = f(u(x), y) \tag{10}$$

for every pair of vectors x, y in E. Replacing x by $u(x)$ in (10) and taking into account the relation $f(u(x), u(y)) = f(x, y)$, we get

$$f(u^2(x) - x, y) = 0 \tag{11}$$

for all x and y, hence $u^2(x) = x$ for all $x \in E$, since f is nondegenerate. In other words, u is an *involution* in $U_n(K, f)$; as K has characteristic 2, this means that $u(x) = x + t(x)$, where t is a linear mapping of E onto a totally isotropic subspace V, $V^* \supset V$ being the kernel of t. In order to prove Theorem 2, we argue as follows: if we can choose a quasi-symmetry v such that uv is not an involution any more, then uv is a product of quasi-symmetries, and so is therefore u. Let a be a vector in V; if $u(v(a)) \neq v^{-1}(u^{-1}(a))$, uv will not be an involution. As $u(a) = a$, and $u(v(a)) = v(a) + b$, where $b \in V$, the relation $u(v(a)) = v^{-1}(u^{-1}(a))$ can be written $v(a) - v^{-1}(a) = b$, and

therefore we have to show that it is possible to find a quasi-symmetry v such that $v(a) - v^{-1}(a)$ is not in V. Take a non-isotropic vector c which is not orthogonal to a; the intersection of the plane $P = aK + cK$ with V is then the line aK. Let v be the quasi-symmetry of hyperplane H orthogonal to c, transforming c into $c\rho$ $(\rho \neq 0)$; let $a = c\lambda + z$, where $z \in H$ and $\lambda \neq 0$; then $v(a) = c\rho\lambda + z$, $v^{-1}(a) = c\rho^{-1}\lambda + z$, hence $v(a) - v^{-1}(a) = c(\rho - \rho^{-1})\lambda$, and this last vector will not be in V provided ρ can be chosen so that $\rho^2 \neq 1$ (i. e. $\rho \neq 1$), and $\rho^J f(c, c)\rho = f(c, c)$. But this is always possible when f is trace-valued and $J \neq 1$, ([4], p. 370), and this ends our proof when $K_0 \neq \mathfrak{F}_2$.

9. It is easy to see that the case $K_0 = \mathfrak{F}_2$, $n = 2$ is exceptional for Theorem 2. In the 2-dimensional vector space over $K = \mathfrak{F}_4$, there are then 5 lines aK through the origin; two of them, e_1K and e_2K, are orthogonal and non isotropic, the other three are isotropic. We have $f(e_1, e_1) = f(e_2, e_2) = 1$, since for a nonisotropic vector x, $f(x, x) \in K_0$ and is not 0, hence must be 1; the linear transformation exchanging e_1 and e_2 belongs therefore to $U_2(\mathfrak{F}_4)$; but it cannot be generated by quasi-symmetries, since these leave invariant each line e_1K, e_2K.

We have still to prove that the theorem is true for $K = \mathfrak{F}_4$ and $n \geq 3$. It will be enough to give the proof for $n = 3$, for if $n > 3$, $u \in U_n(\mathfrak{F}_4)$, and x is a nonisotropic vector in E, the two vectors x and $u(x)$ belong to a nonisotropic 3-dimensional subspace V: this is immediate if $u(x) = x$, or if the plane P through x and $u(x)$ is non isotropic. On the other hand, if P is isotropic, it is not totally isotropic, since it contains the nonisotropic vector x; there is therefore only one isotropic line aK in P, orthogonal to P; if b is a nonisotropic vector in E, not orthogonal to a, the 3-dimensional subspace $V = P + bK$ is nonisotropic, for if we write that $y = a\alpha + b\beta + x\gamma$ is orthogonal to a, b and x, we find $f(a, b)\beta = 0$, $f(b, a)\alpha + f(b, b)\beta + f(b, x) = 0$ and $f(x, b)\beta + f(x, x)\gamma = 0$; as $f(a, b) \neq 0$, $\beta = 0$, and as $f(x, x) \neq 0$, $\gamma = 0$; then the second equation reduces to $f(b, a)\alpha = 0$, which shows that $y = 0$ and proves our contention. The theorem being supposed to be true for $n = 3$, there exists *in* V a product of quasi-symmetries transforming x into $u(x)$ (since by Witt's theorem there is a unitary transformation in V transforming x into $u(x)$); these quasi-symmetries in V can be extended to quasi-symmetries *in* E by taking them to be the identity in the orthogonal subspace V^*. Then, the inductive process used in section 8 can again be applied for $n > 3$, and the proof is thus reduced to the case $n = 3$.

10. To prove the theorem in that case, we will first show that if a and b are any two nonisotropic vectors in E, there is a product of quasi-symmetries

transforming a into b. This result is obvious if $b = a\alpha$ with $\alpha \neq 0$ (and then necessarily $\alpha\alpha^J = 1$, an equation which is verified by every element $\alpha \neq 0$ in $\mathfrak{F}_4$); for then we merely consider the quasi-symmetry s of hyperplane orthogonal to a, and such that $s(a) = a\alpha = b$. Suppose next a and b are not collinear and consider an orthogonal basis e_1, e_2, e_3 of E such that $e_1 = a$; the only nonisotropic lines orthogonal to e_1K are e_2K and e_3K; the other 9 nonisotropic lines in E are in the 3 isotropic planes through e_1K and the 3 isotropic lines in the plane $e_2K + e_3K$ (3 such lines distinct from e_1K in each of these planes). Suppose first that b is not orthogonal to a; then bK is in one of the isotropic planes through e_1K, say P. Let cK be one of the nonisotropic lines in P distinct from aK and bK; as there are 3 elements $\alpha \in \mathfrak{F}_4$ such that $\alpha\alpha^J = 1$, there are 3 distinct quasi-symmetries of hyperplane H orthogonal to cK, transforming c into the 3 vectors $c\alpha$, and leaving elementwise invariant the (unique) isotropic line D in P, orthogonal to c. As such a quasi-symmetry cannot leave invariant any line in P other than cK and D, the 3 preceding quasi-symmetries transform aK into the 3 lines in P distinct from cK and D; in particular there is one such quasi-symmetry transforming aK into bK, and we have then proved our assertion.

11. There remains the case in which a and b are orthogonal; we can suppose for instance that $a = e_1$, $b = e_2$. We are going to prove slightly more, namely that there is a product of quasi-symmetries which *exchanges* e_1K *and* e_2K. Let D be an isotropic line in the plane $e_1K + e_2K$, and let cK be a nonisotropic line distinct from e_3K in the isotropic plane $P = D + e_3K$. As cK is not orthogonal to e_1K, the plane $Q = e_1K + cK$ is isotropic; there is therefore a quasi-symmetry s of hyperplane orthogonal to c, which will transform the line e_1K into another nonisotropic line xK in the plane Q. This line xK is neither in the plane $e_1K + e_2K$ (for the intersection of Q with that plane is e_1K), nor in the plane $e_3K + e_2K$ (for the intersection of that plane with Q is isotropic). Therefore the plane $R = e_2K + xK$ is isotropic. The intersection $R \cap P = dK$ is then nonisotropic, since the only isotropic line in P is D in the plane $e_1K + e_2K$; and the argument in section 10 shows that there is a quasi-symmetry s' of hyperplane orthogonal to dK, which will transform xK into e_2K. This shows that $s's$ transforms e_1K into e_2K; but in addition, both s and s' leave invariant the isotropic line D, hence $s's$ transforms the plane $e_1K + e_2K = e_1K + D = D + e_2K$ into itself; but it then transforms e_2K, which is orthogonal to e_1K in the plane $e_1K + e_2K$, into the line orthogonal to e_2K in that plane, that is e_1K.

We can now end the proof in the following way: if x is a nonisotropic vector in E, there is a product v of quasi-symmetries which transforms $u(x)$ into x; $vu = u_1$ then leaves invariant x, hence also the plane H orthogonal to x in E; let aK and bK be the nonisotropic lines (orthogonal to each other) in H. If u_1 leaves invariant both these lines, it is a product of quasi-symmetries; if it exchanges aK and bK, then there is a product w of quasi-symmetries in E which exchanges aK and bK, and therefore leaves invariant xK; the product wu_1 then leaves invariant xK, aK and bK, and is therefore a product of quasi-symmetries. Theorem 2 is thus completely proved.

12. The assumptions in this section are those of section 8, with the additional condition that $J \neq 1$; the orthogonal and symplectic groups are thus excluded. Let ϕ be the natural homomorphism of K^* onto its factor group K^*/C, where C is the commutator subgroup of K^*.

THEOREM 3. *For every unitary transformation $u \in U_n(K, f)$, the determinant* [1] *of u has the form $\phi(\gamma^J\gamma^{-1})$.*

Theorem 2 reduces the proof of Theorem 3 to the case in which u is a quasi-symmetry (the exceptional case of Theorem 2 is of course known to satisfy also Theorem 3, which is well known for commutative fields K). Let us therefore suppose that u is a quasi-symmetry of hyperplane H, and let e be a vector orthogonal to H, and $\rho = f(e, e) \neq 0$; then $u(e) = e\alpha$ with $\alpha^J\rho\alpha = \rho$, and the determinant of u is obviously $\phi(\alpha)$ (take as a basis of E the vector e and a basis of H). We are thus reduced to prove that the relation $\alpha^J\rho\alpha = \rho$ implies $\alpha = \gamma^J\gamma^{-1}$ for some $\gamma \neq 0$. This is immediate if $\alpha = -1$, for if λ is an element in K such that $\lambda^J \neq \lambda$, then for $\gamma = \lambda - \lambda^J$, $\gamma^J = -\gamma$, and as $\gamma \neq 0$, $-1 = \gamma^J\gamma^{-1}$. If $\alpha \neq -1$, consider the element $\gamma^{-1} = \rho(1 + \alpha)$; we have $\gamma^{-J} = (1 + \alpha^J)\rho$, hence $\gamma^{-J}\alpha = \rho\alpha + \alpha^J\rho\alpha = \rho\alpha + \rho = \gamma^{-1}$, whence $\alpha = \gamma^J\gamma^{-1}$, and this ends our proof.

13. We add two disconnected remarks on results proved in [4] and [2] respectively. The first concerns Theorem 4 of [4], p. 380 which is proved in that paper for a sfield K of characteristic $\neq 2$. I want to show that the theorem is still valid when that restriction on the characteristic is dropped. To prove this, let us first remark that we can suppose that there is a vector $a \in E$ such that $f(a, a) = 1$ (excluding the symplectic groups, whose structure is well known). Indeed, let a be a nonisotropic vector, and let $f(a, a) = \alpha \neq 0$; α is a symmetric element for the involution J, the form $f_1(x, y) = \alpha^{-1}f(x, y)$ is hermitian for the involution $\xi \to \xi^T = \alpha^{-1}\xi^J\alpha$; moreover, f_1 is trace-valued

for that involution, for we have $\alpha^{-1}(\xi + \xi^J) = (\alpha^{-1}\xi) + (\alpha^{-1}\xi)^T$; finally, $f_1(a, a) = 1$, and it is clear that the unitary groups $U_n(K, f)$ and $U_n(K, f_1)$ are identical. With this assumption, we have therefore $1 = \rho + \rho^J$; Theorem 4 of [4] will then be proved, if we can prove Lemma 6, [4], p. 380, and this in turn amounts to finding 4 elements α, β, α_1, β_1 in K such that $\alpha_1\alpha_1^J = \alpha\alpha^J$, $\beta_1\beta_1^J = \beta\beta^J$, and $\alpha_1\beta_1^J - \alpha\beta^J = \gamma$, where γ is an arbitrarily given element of K. In order to show that this is possible, take $\alpha_1 = \alpha$, and $\beta_1 = 1$; if we can prove that β can be chosen such that $\beta\beta^J = 1$ and $\beta \neq 1$, the equation $\alpha_1\beta_1^J - \alpha\beta^J = \gamma$ reduces to $\alpha(1 - \beta^J) = \gamma$ and always has a solution. But if we take $\beta = \rho^J\rho^{-1}$, then $\beta\rho\beta^J = \rho$, $\beta\rho^J\beta^J = \rho^J$, hence $\beta\beta^J = \beta(\rho + \rho^J)\beta^J = \rho + \rho^J = 1$, and $\rho^J \neq \rho$, since otherwise $\rho + \rho^J$ would be 0.

14. The other remark is relative to the proof, given in [2], pp. 72-73, that every noncommutative sfield K for which every element has degree ≤ 2 over the center Z of K, is a reflexive sfield. The first remark in the proof is that every finite set $(\xi_i)_{1 \leq i \leq n}$ of elements of K generates over Z a sfield $Z[\xi_1, \cdots, \xi_n]$ of rank $\leq 2^n$; but this does not (contrary to what is asserted in [2], p. 73) reduce the problem to the case in which $[K : Z]$ is finite, for the center of $K[\xi_1, \cdots, \xi_n]$ might be distinct from Z. To fill that gap in the proof, let us consider two nonpermutable elements ξ, η of K; as the rank $[Z[\xi, \eta] : Z]$ is ≤ 4 and is a multiple of 2 which cannot be equal to 2 (otherwise $Z[\xi, \eta]$ would be commutative) it is equal to 4. Let now ζ be any other element of K, and let T be the center of $Z[\xi, \eta, \zeta]$; we want to prove that $T = Z$, for then the argument of ([2], p. 73) shows that $Z[\xi, \eta, \zeta]$ has rank 4 over Z, hence $\zeta \in Z[\xi, \eta]$, and therefore $K = Z[\xi, \eta]$. Suppose $T \neq Z$, hence as $[Z[\xi, \eta, \zeta] : Z] \leq 8$, we have $[Z[\xi, \eta, \zeta] : T] < 8$, and as the rank of $Z[\xi, \eta, \zeta]$ over its center T must be a square and cannot be 1, $[Z[\xi, \eta, \zeta] : T] = 4$, and as $T \neq Z$, $[T : Z] = 2$. There is then a maximal subfield S of $Z[\xi, \eta, \zeta]$ such that $[S : T] = 2$ and that S is separable over T. We can write therefore $S = T(\alpha)$, where α is separable and of degree 2 over T; but α is also of degree 2 over $Z \subset T$, hence is separable over Z. On the other hand, we have $T = Z(\beta)$, where β has degree 2 over Z, but might be inseparable when Z has characteristic 2. Nevertheless, in such circumstances, it is known ([5], p. 132) that for the field $S = Z(\alpha, \beta)$, the theorem of the primitive element still holds, in other words $S = Z(\gamma)$, where γ has degree 4 over Z; but this contradicts the assumption that the degree of γ over Z is at most 2, and the assumption $T \neq Z$ is therefore untenable, which ends the proof.

UNIVERSITY OF MICHIGAN.

BIBLIOGRAPHY.

[1] J. Dieudonné, "Les déterminants sur un corps non commutatif," *Bulletin de la Société Mathématique de France*, vol. 71 (1943), pp. 27-45.

[2] ———, "Sur les groupes classiques," *Actualités Scientifiques et industrielles*, no. 1040, Paris (Hermann), 1948.

[3] ———, "On the automorphisms of the classical groups," *Memoirs of the American Mathematical Society*, no. 2 (1951).

[4] ———, "On the structure of unitary groups," *Transactions of the American Mathematical Society*, vol. 72 (1952), pp. 367-385.

[5] B. L. van der Waerden, *Moderne Algebra*, t. I (2nd ed.); Berlin (Springer), 1937.

ON THE LOCAL RÔLE OF THE THEORY OF THE LOGARITHMIC POTENTIAL IN DIFFERENTIAL GEOMETRY.*

By AUREL WINTNER.

Definitions. In a (u, v)-plane, let D be an open domain (which, for the purposes at hand, can be assumed to be simply connected and "sufficiently small") and let $g_{11}, g_{12} = g_{21}, g_{22}$ be three functions which are of class C^n and satisfy the conditions $g_{ii} > 0$, $\det g_{ik} > 0$ on D. Then

$$g_{\alpha\beta}(u^1, u^2)du^\alpha du^\beta, \quad \text{where} \quad u^1 = u, u^1 = v, \tag{1}$$

will be called a C^n-metric (on D).

Let

$$g'_{\alpha\beta}(u'^1, u'^2)du'^\alpha du'^\beta, \quad \text{where} \quad u'^1 = u', u'^2 = v', \tag{2}$$

be a C^m-metric on a (u', v')-domain D'. Then the C^n-metric (1) is called isometric to (2) if, corresponding to every point (u_0, v_0) of D, there exist a circle

$$D_0\colon\ (u - u_0)^2 + (v - v_0)^2 < \epsilon^2 \tag{3}$$

and a C^1-mapping

$$u' = u'(u, v), \qquad v' = v'(u, v) \tag{4}$$

of (3) on a (*schlicht*) domain D'_0 in D' in such a way that (1) becomes identical with (2) on D_0 (or D'_0) by virtue of (4) (or

$$u = u(u', v'), \qquad v = v(u', v'), \tag{5}$$

the inverse of (4)). It is understood that (4) is called a (local) C^r-mapping, where $r \geqq 1$, if both functions (4) are of class C^r and have a non-vanishing Jacobian (which implies that, if (3) is small enough, D'_0 is *schlicht* and (5) is a C^r-mapping of D'_0 onto D_0).

If (1) is a C^n-metric, where $n \geqq 2$, then it has a curvature $K = K(u, v)$, defined, if

$$g = (\det g_{ik})^{\frac{1}{2}} \qquad (g > 0), \tag{6}$$

---

* Received April 25, 1953.

by

$$(7) \qquad -2gK = \{(g_{22u} - g_{12v})/g\}_u + \{(g_{11v} - g_{12u})/g\}_v + (2g^3)^{-1} \det(\Gamma, \Gamma_u, \Gamma_v), \quad \text{where} \quad \Gamma = (g_{11}, g_{12}, g_{22})$$

(the subscripts u and v denote partial differentiations). Thus $K(u, v)$ is a function of class C^{n-2} if (1) is a C^n-metric, where $n \geqq 2$.

In particular, (1) must be a C^3-metric before it is *sure* to have a $K(u, v)$ of class C^1. On the other hand, $K(u, v)$ *can* have this property also if (1) is just a C^2-metric. Actually, it turns out that this type of a metric, a C^2-metric (1) for which the curvature $K(u, v)$ is a function of class C^1 (in terms of the parameters (u, v) in which the coefficients $g_{ik}(u, v)$ of (1) are given as functions of class C^2), is quite an important class. In fact, it will be shown that for *this* class of metrics there prevail simple facts (two of which, Theorem 1 and Theorem 3, become false if they are referred to either of the, more general or less general, classes of C^n-metrics, where $n = 2, 3$).

Conformal normal forms. Let a metric (1) be called *conformal* (that is to mean, "conformal with reference to the euclidean (u, v)-plane," i. e., "isothermic") if

$$(8) \qquad g_{11} = g_{22} \quad \text{and} \quad g_{12} = 0$$

hold identically in (u, v); so that (1) reduces to

$$(9) \qquad g(u, v)(du^2 + dv^2), \qquad (g > 0),$$

if g denotes the common value of g_{11} and g_{22} (in view of (8), this agrees with the more general notation (6)). According to Lichtenstein [4], every C^1-metric (1) is isometric to a conformal metric

$$(10) \qquad g'(u', v')(du'^2 + dv'^2) \qquad (g' > 0).$$

In fact, he proved that this is true also if the coefficients $g_{ik}(u, v)$ of (1), instead of being functions of class C^1, are only of class $C^{\epsilon-0}$ (for some $\epsilon > 0$), where $C^{\lambda-0}$ denotes (for $0 < \lambda \leqq 1$) the class of those functions which belong for every $\mu < \lambda$ to the class C^μ of those functions which satisfy a locally uniform Hölder condition of index μ. On the other hand, it was shown in [2] that not every metric (1) with continuous coefficients $g_{ik}(u, v)$ is isometric to a (continuous) conformal metric.

While Lichtenstein's method [4] also proves that every C^2-metric is isometric to a conformal C^1-metric, it was shown in [2] that not every

C^2-metric is isometric to a conformal C^2-metric. In order that a metric be isometric to a conformal C^2-metric, it is sufficient to assume that the former is a C^3-metric but it is not true that every C^3-metric is isometric to a conformal C^3-metric; cf. [4] and [2], respectively. It will be shown that this disorder can be disposed of by considering the class of C^2-metrics possessing a curvature of class C^1, a class intermediary between the class of the C^2-metrics and that of the C^3-metrics.

THEOREM 1. *Every C^2-metric* (1) *possessing a curvature $K(u, v)$ of class C^1 is isometric to a conformal C^2-metric* (10) *possessing a curvature $K'(u', v')$ of class C^1 (and every C^1-mapping* (4) *establishing the isometry must be a C^3-mapping*).

Accordingly, every C^2-metric possessing a curvature $K(u, v)$ of class C^1 (but, as mentioned before, not an arbitrary C^2-metric, the curvature of which is just continuous in general) must be isometric to a conformal C^2-metric. Actually, the latter statement is not weaker than Theorem 1, since, the curvature

$$K'(u', v') = K(u, v) = K(u(u', v'), v(u', v')) \tag{11}$$

being a scalar, its C^1-character is preserved by every C^1-mapping (4). Correspondingly, the parenthetical assertion of Theorem 1 is not deeper than (and is of course contained in) in the *second* of the following statements (I), (II), $\cdots$:

(I) If (1) and (10) are isometric C^1-metrics, then every C^1-mapping (4) establishing their isometry must be a C^2-mapping.

(II) The assertion of (I) remains true if C^1 and C^2 are replaced by C^2 and C^3, respectively. Etc.

(I), (II), $\cdots$ can be proved in various ways. The significance of these and of more general conclusions was pointed out in [8], pp. 199-203. The most general result in the binary and definite case was proved by Hartman [1].

The simplest proof of (I), (II), $\cdots$ seems to be a reduction to the particular case in which (10) is the euclidean metric ($g' \equiv 1$). Let, (I*), (II*), $\cdots$ denote the particular statements which thus result from (I), (II), $\cdots$, respectively (that is, the statements in which

$$du'^2 + dv'^2 \tag{12}$$

replaces (10) in (I), (II), $\cdots$). For a proof of (I*), which is valid for any dimension number and for indefinite metrics also, cf. [7], pp. 571-573.

The proof of (II*) is the same as that of (I*) (but in one respect simpler; cf. *loc. cit.*). But the particular cases (I*), (II*) of (I), (II) imply the latter. This can be seen as follows: Under the assumptions and in the notations of (I), define a (positive) function f of (u, v) by placing

$$f(u, v) = 1/g'(u'(u, v), v'(u, v)). \tag{13}$$

Then $f(u, v)$ is of class C^1, since $g'(u', v')$ and both functions (4) are. According to (13), the isometry of (10) and (1) means the isometry of the euclidean metric (12) and of the C^1-metric

$$f(u^1, u^2) g_{\alpha\beta}(u^1, u^2) du^\alpha du^\beta, \quad \text{where} \quad u^1 = u, \; u^2 = v. \tag{14}$$

Hence, the assertion of (I) follows by applying (I*) to (14) and (12). Similarly, (II) follows from (II*) if use is made of (I).

There remains to be proved the main assertion of Theorem 1. The proof will be such as to supply the following results as a by-product:

THEOREM 2. *If the curvature $K(u, v)$ of a C^2-metric* (1) *is a function of class C^j, and if $j \leqq 3$, then* (1) *is isometric to a C^{j+1}-metric.*

Incidentally, the latter metric will result as a *conformal* C^{j+1}-metric.

One might expect that some information on the C^2-metric will also result if the C^j-character of $K(u, v)$ is required for an index $j \geqq 4$. The place assigned to C^2-metrics (1) by the scale (I), (II), $\cdots$ seems, however, to indicate that the process of refining is arrested at $j = 3$.

That scale prevents, in particular, the extension of Theorem 2 to a scale the next item of which would assert that, if $j = 4$, the C^2-metric must be isometric to a C^5-metric (but a counterexample is missing).

Proof of Theorem 1. Let (1) be a C^2-metric on a (u, v)-domain D. Then the function K of (u, v) is just continuous, while the functions (6) and Γ, occurring in (7), are of class C^2 and C^1, respectively. Hence, if $B = B(J)$ denotes the interior of a positively oriented, rectifiable Jordan curve J contained in D, then an integration of (7) over B, when followed by an application of Green's formula, leads to the following identity in $J = J(B)$:

$$\int_J (2g)^{-1}\{g_{11v} - g_{12u}) du + (g_{12v} - g_{22u})\} dv \tag{15}$$

$$= \iint_B g\{K + (4g^4)^{-1} \det(\Gamma, \Gamma_u, \Gamma_v)\} du dv, \quad \text{where} \quad \Gamma = (g_{11}, g_{12}, g_{22}).$$

Actually, the existence of a continuous function $K = K(u, v)$ satisfying this integral identity can be thought of as defining a C^1-metric (1) which possesses a continuous curvature (note that (7) does not apply to a metric which is just of class C^1). This idea is due to Weyl [5], pp. 42-44.

Suppose that the C^2-metric (1) occurring in (15) is transformed by a C^1-mapping (4) into a metric (2) which, for some reason, is known to be a C^1-metric (rather than just a continuous metric) on the (u', v')-image D' of D. Then, according to [1], p. 222, the C^1-mapping (4) must be a C^2-mapping. Denote by (15′) the relation which results if $u, v, g_{ik}, g, \Gamma, K$ (and J, B) are replaced by $u', v', g'_{ik}, g', \Gamma', K'$ (and the (u', v')-images J', B' of J, B), respectively. Then it can be verified from the definitions $g = (\det g_{ik})^{\frac{1}{2}}$, $g' = (\det g'_{ik})^{\frac{1}{2}}$ and from the tensor character of (g_{ik}), (g'_{ik}) that, if (11) is considered as the definition of $K'(u', v')$, the relation (15′) holds as an identity in J' or B', where $B' = B'(J')$ denotes the interior of any positively oriented, rectifiable Jordan curve contained in D'. It follows (if (1) and (2) are interchanged) that the identity (15) in J holds not only for C^2-metrics but also for every C^1-metric which (for some reason) is isometric to a C^2-metric.

As explained after Theorem 1, the latter is equivalent to the statement that every C^2-metric (2) possessing a curvature $K'(u', v')$ of class C^1 is isometric to a conformal C^2-metric (2), that is, to a metric (9) in which $g(u, v)$ is a function of class C^2. In order to prove the latter formulation of Theorem 1, assume first only that the given metric (2) is a C^2-metric. According to a theorem of Lichtenstein, mentioned above, any such metric (2) is isometric to a conformal C^1-metric; that is, to a metric (9) in which $g(u, v)$ is a function of class C^1. Since the C^1-metric (9) is isometric to a C^2-metric, (15) is applicable to (9) and reduces, in view of (8), to

$$\int_J (2g)^{-1}(g_v du - g_u dv) = \iint_B gK du dv. \tag{16}$$

Hence, in order to prove Theorem 1, it remains to be shown that if not only $g(u, v)$ but also $K = K(u, v)$ (cf. (11) and (4)) is of class C^1, an assumption of Theorem 1 which was not used thus far, then (16) implies that $g(u, v)$ is of class C^2.

To this end, note that if

$$f(u, v) = -2K(u, v)g(u, v) \tag{17}$$

and

$$\gamma(u, v) = \log g(u, v), \qquad (g > 0), \tag{18}$$

then (16) can be written in the form

$$\int_J (\gamma_u dv - \gamma_v du) = \iint_B f(u,v)\,du\,dv \tag{19}$$

and that (19) is just the integrated form of Poisson's equation

$$\gamma_{uu} + \gamma_{vv} = f(u,v). \tag{20}$$

Correspondingly, the proof of Theorem 1 must be made to depend on an appeal to the theory of the logarithmic potential

$$\phi(u,v) = \tfrac{1}{4\pi}\iint_A f(x,y)\log\{(u-x)^2 + (v-x)^2\}\,dx\,dy. \tag{21}$$

It is known that if $f(u,v)$ is an arbitrary continuous function (not satisfying anything like a Hölder condition), then (i) the differential equation (20) need not possess any (continuous) solution $\gamma(u,v)$ but (ii) the integrated form (19) of (20) is satisfied by the sum, $\gamma(u,v)$, of any regular harmonic function $h(u,v)$ and the logarithmic potential (21), which is a function $\phi(u,v)$ of class C^1, finally (iii) every function γ of class C^1 satisfying (19) is of the form $h+\phi$; cf. [9], [3], [6]. It is understood that the continuous function $f(u,v)$ is supposed to be given on a (u,v)-domain D and that, if A is any bounded domain the closure of which is contained in D, then, in the above assertions, (21) is considered only for the domain of points (u,v) which constitute A.

Since $g(u,v)$ and (by assumption) $K(u,v)$ are functions of class C^1, the same is true of the function $f(u,v)$ defined by (17). Hence the corresponding logarithmic potential (21) is a function of class C^2 on A (the conclusion that (21) is of class C^2 on A, whenever $f(u,v)$ is of class C^1, is due to Gauss; the finer aspects of Hölder's theory are not needed). Since every function γ of class C^1 satisfying (19) is of the form $h+\phi$, it follows that every such γ is of class C^2. Finally, since (18) is a (particular) function of class C^1 satisfying (19), it follows that the γ occurring in (18), and therefore $g = e^\gamma$ as well, is of class C^2.

This proves Theorem 1.

Proof of Theorem 2. If $j=1$, then Theorem 2 and the remark made after it are contained in Theorem 1. The remaining two cases, $j=2$ and $j=3$, must be dealt with in succession.

Suppose first that the assumptions of the case $j = 2$ of Theorem 2 are satisfied. Then, in view of Theorem 1 (including its parenthetical assertion; cf. (11)), it can be assumed that the given metric is of the type (9). Thus both g and K are now of class C^2, and the assertion is that g must therefore be of class C^3.

Corresponding to the result of Gauss, used above, the logarithmic potential (21) of every density f of class C^2 is of class C^3. Since the product (17) is of class C^2, it follows, by the argument applied at the end of the proof of Theorem 1, that g is of class C^3.

This proves Theorem 2, as well as the remark made after Theorem 2, for the case $j = 2$. The corresponding assertion for the case $j = 3$ follows in the same way as in the case $j = 2$ (if use is made of results proved for the latter case).

General indices. All of this can be summarized as the case $n = 2$ of the following theorem (*):

(*) *If* (1) *is a* C^n*-metric, where* $n > 1$, *and if* (1) *has a curvature* $K(u, v)$ *of class* C^{n-1+m}, *where* $0 \leqq m \leqq 2$, *then* (1) *is isometric to a conformal* C^{n+m}*-metric* (10) (*and* K *is a function of class* C^{n-1+m} *in terms of the parameters* u', v' *of* (10) *also*).

Clearly, the above proofs, given for $n = 2$, hold for $n = 3, 4, \cdots$ also. The necessity of restricting m to its three lowest values remains undecided (for every n); cf. the remarks made after Theorem 2.

With regard to the case $n = 1$, excluded in the wording of (*), it is clear from the proof of Theorem 2 that if (*) is true for $(n, m) = (1, 0)$, then it is true for $(n, m) = (1, 1)$ and $(n, m) = (1, 2)$ also (here a C^1-metric (1) possessing a $K(u, v)$ of class C^0 is defined in the sense of Weyl, as explained after (15)). But the truth of (*) for the case $(n, m) = (1, 0)$ remains undecided.[1]

---

[1] Weyl has claimed ([5], p. 49) that the truth of (*) for $(n, m) = (1, 0)$ is known from the theory of functions (in his case of a closed convex surface, where $K > 0$). Actually, all that follows by function-theoretical methods is that if the theorem is true in the small, then (subject to trivial topological restrictions) it is true in the large also. The sharpest known results in the small seem to be those of Lichtenstein, referred to above.

An apparent exception seems to be contained between the lines of a proof of a (non-local) theorem of Weyl [5], p. 68, in which the metric (1), instead of satisfying the assumptions of Theorem 1, is merely required to be a C^1-metric possessing a curvature $K(u, v)$ which, along with the first derivatives of the coefficients of (1), is sub-

Surfaces. By a surface of class C^n, where $n \geqq 1$, will be meant a set S of points $X = (x, y, z)$ having the property that, in a neighborhood of every point of it, S can be parametrized in the form

$$X = X(u, v), \qquad (X = (x, y, z)), \tag{22}$$

where (u, v) varies on some, sufficiently small, domain D, the vector function (22) is of class C^n and the vector product, say $V = V(u, v)$, of the partial derivatives X_u, X_v does not vanish; so that $N = V/|V|$ defines a unit vector which is a function $N(u, v)$ of class C^{n-1}. Then (22) will be called a C^n-parametrization of the surface S of class C^n. Note that, if $n > m \geqq 1$, every surface S of class C^n possesses C^m-parametrizations which are not C^{m+1}-parametrizations.

If (22) is of class C^3, then, since $V \neq 0$, the first fundamental form

$$|dX(u^1, u^2)|^2 = g_{\alpha\beta}(u^1, u^2)\, du^\alpha du^\beta \qquad (u^1 = u, u^2 = v) \tag{23}$$

on S is a metric (1) of class C^2. But more than this is true:

THEOREM 3. *In order that a C^2-metric* (1) *be realizable as the metric* (23) *on some surface* (22) *of class C^3, it is necessary for* (1) *to have a curvature $K(u, v)$ of class C^1.*

In fact, if $n \geqq 2$, then, since the unit vector $N(u, v)$ is of class C^{n-1}, the second fundamental form on S, being defined as the scalar product of $-dN(u, v)$ and $dX(u, v)$, exists and has coefficients, say $h_{ik} = h_{ki}$, which are functions of class C^{n-2}. Since the coefficients of (23) are functions of class C^{n-1} and since, due to the embedding (22)-(23) of the metric (1) on S, the ratio of the determinants $\det h_{ik}(u, v)$, $\det g_{ik}(u, v) > 0$ is identical with the curvature $K(u, v)$ of the metric (Gauss), it follows that $K(u, v)$ is of class C^{n-1} if $n > 2$, and exists and is continuous if $n = 2$ (Weyl; cf. (15)). Theorem 3 is the case $n = 3$ of this conclusion, which applies the assertion of the *theorema egregium* in the direction just opposite to that emphasized by Gauss.

Let a C^n-parametrization (22), where $n \geqq 1$, be called an *isothermic parametrization of a surface S* if (8) holds for the first fundamental form (23). Then it follows from Theorem 3 that Theorem 1 (along with the

---

jected to a Hölder condition (*loc. cit.*, K is positive on a closed orientable (u, v)-manifold). It turns out, however, that any such metric must be isometric to a C^2-metric and, what is more, to a Hölderian C^2-metric. This is the first assertion of Theorem 5 below; cf. also Theorem 6.

parenthetical assertion of Theorem 1) contains the following fact as a particular case:

THEOREM 4. *Every surface S of class C^3 possesses isothermic C^3-parametrizations.*

The above proofs make it clear that Theorem 4 remains true if both of its classes C^3 are replaced by C^n, provided that $n > 2$ (and it seems to be likely that $n = 2$ can here be included), whereas Lichtenstein's theory [4] supplies only the existence of an isothermic parametrization of class C^{n-1} if the surface S is of class C^n, where $n = 2, 3, \cdots$. This shows the "natural" character of the assumptions and assertions of Theorem 1 (and of its analogues for other n-values).

Hölder restrictions. Let a function $f(u, v)$, defined on a (u, v)-domain D, be called of class HC^n, where $n \geqq 0$, if it is of class C^n and all of its partial derivatives $\partial^{h+j} f / \partial u^h \partial v^j$, where $h + j \leqq n$, satisfy a (locally) uniform Hölder condition of *some* positive index $\lambda (= \epsilon < 1)$. Correspondingly, a mapping (4) will be called an HC^n-mapping, where $n \geqq 1$, if it is a C^n-mapping (with non-vanishing Jacobian) and both functions u', v' occurring in (4) are of class HC^n. Similarly, a metric (1) will be called an HC^n-metric, where $n \geqq 0$, if all three functions $g_{ik}(u, v)$ are of class HC^n.

In this terminology, a variant of Theorem 1 can be formulated as follows:

THEOREM 5. *Let* (1) *be an HC^1-metric possessing a curvature $K(u, v)$ of class HC^0. Then* (1) *is isometric to an HC^2-metric. The latter can be chosen to be a conformal metric. Finally, the mapping* (4)-(5), *establishing the isometry of the given metric* (1) *and the conformal HC^2-metric* (10), *is realized by an HC^2-mapping.*

The same is true if HC^1, HC^0, HC^2 are replaced by HC^n, HC^{n-1}, HC^{n+1}, respectively, where $n \geqq 1$.

The implications of this theorem are revealed by the following fact:

THEOREM 6. *Every conformal C^n-metric possessing a curvature of class HC^{n-1} is an HC^{n+1}-metric, where $n \geqq 1$.*

It can readily be concluded from the results of Lichtenstein, quoted before Theorem 1, that the case $n = k \geqq 1$ of Theorem 5 follows if Theorem 6 is proved for the case $n = k + 1 \geqq 2$. But Theorem 6 holds for every $n \geqq 1$.

In order to see this, let $n = 1$ (if $n > 1$, the proof is the same, though more straightforward). Then the assumption of Theorem 6 is that a function $g(u, v) > 0$ of class C^1 and a function $K(u, v)$ of class HC^0 satisfy (16) as an identity in $B = B(J)$. In view of (17) and (18), this means that $\gamma(u, v)$ is a function of class C^1 satisfying (19), where $f(u, v)$ is a function subject to a locally uniform Hölder condition. Since (19) is the integrated form of (20), it now follows (for the reasons (i)-(iii), mentioned after (20)), that $\gamma(u, v)$ is of class C^2 and that its second derivatives satisfy a locally uniform Hölder condition. It follows therefore from (18) that $g(u, v)$ is of class HC^2, as claimed by the case $n = 1$ of Theorem 6.

Appendix.*

Theorem 3 and its extensions to the classes $C^4, C^5, \cdots$ can be formulated as follows: *If a surface S has a C^n-parametrization* (22), *where $n > 2$, then the C^{n-1}-metric* (23) *is isometric to a conformal C^{n-1}-metric.* It is natural to ask whether there is a corresponding theorem for the case in which the first fundamental form (23) is replaced by the second fundamental form

$$-dN(u, v) \cdot dX(u, v) = h_{\alpha\beta}(u^1, u^2)\, du^\alpha du^\beta, \tag{24}$$

provided that (24) is a metric, i. e., provided that the curvature $K(u, v)$ of (22) or (23) is positive. The answer to this question turns out to be positive: *If a surface S of postive curvature K has a C^n-parametrization* (22), *where $n > 3$, then the C^{n-2}-metric* (24) *is isometric to a conformal C^{n-2}-metric.* This follows from Theorem 1 and its extensions to C^k-metrics, where $k \geqq 2$, in the same way as the theorem italicized before (24) does, if the following counterpart of Theorem 3 is proved:

THEOREM 3 bis. *If a surface S of non-vanishing curvature K has a C^n-parametrization* (22), *where $n > 2$, then, although the coefficients $h_{ik}(u, v)$ of* (24), *where $(u, v) = (u^1, u^2)$, are just of class C^{n-2} in general, the binary quadratic differential form* (24) (*which is definite or indefinite according as the curvature $K(u, v)$ of* (23) *is positive or negative*) *possesses a curvature $\kappa(u, v)$ of class C^{n-3}* (which means the existence of continuous $\kappa(u, v)$ in the limiting case $n = 3$).

In the proof of the theorem italicized after (24), only the case $K(u, v) > 0$ of Theorem 3 bis is needed (and only for $n > 3$). At points (u, v) at which $K(u, v)$ vanishes, the assertion of Theorem 3 bis cannot even

* Added July 28, 1953.

be formulated, since $\kappa(u, v)$ is undefined unless either $K(u, v) > 0$ or $K(u, v) < 0$.

Proof of Theorem 3 bis. Suppose first that $n > 3$, say $n = 4$. Then the coefficients $g_{ik}(u, v)$ and $h_{ik}(u, v)$ of (22) and (23) are of class C^3 and C^2, respectively. Hence both $h_{ik}(u, v)$ and the Christoffel coefficients $\Gamma_{ik}^{j}(u, v)$ of (22) are functions of class C^2. But the equations of Codazzi-Mainardi state that the differences $h_{11v} - h_{12u}$, $h_{22u} - h_{12v}$ are certain bilinear forms in h_{ik}, Γ_{ik}^{j}. Consequently, these differences are functions of class C^2 (in (u, v)), which implies that the derivatives

$$(h_{11v} - h_{12u})_v, \quad (h_{22u} - h_{12v})_u \tag{25}$$

are functions of class C^1.

On the other hand, the representation (7) of the curvature $K(u, v)$ of the definite metric (22) shows that the principal part of $K(u, v)$, the part of $K(u, v)$ in which only the contributions of the second derivatives of the functions $g_{ik}(u, v)$ are retained, is precisely the sum of the two expressions which result if the metric $\| h_{ik} \|$ is replaced by $\| g_{ik} \|$ in (25). Correspondingly, the principal part of the curvature $\kappa(u, v)$ of (24) is the sum of the two functions (25), provided that (24) is definite ($\det h_{ik} > 0$, i. e., $K > 0$). Hence, under this proviso, the principal part of $\kappa(u, v)$, and therefore $\kappa(u, v)$ itself, will be of class C^1, since both functions (25) are.

This proves Theorem 3 bis for the case $n = 4$ if $\det h_{ik} > 0$. If $\det h_{ik} < 0$, the proof remains the same, except that the definition of the curvature $\kappa(u, v)$ of an indefinite binary metric must then be used. It is also clear that, in both cases, the proof remains the same if $n = 4$ is replaced by any $n > 3$. Finally, the proof remains valid in the limiting case $n = 3$ also, except that the equations of Codazzi-Mainardi and the definition of the curvature of a non-singular binary metric must then be replaced by their integrated forms (forms which correspond to the replacement of (7), (20) by (15), (19), respectively).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "On unsmooth two-dimensional Riemannian metrics," *American Journal of Mathematics*, vol. 74 (1952), pp. 215-226.

[2] ——— and A. Wintner, "On the existence of Riemannian manifolds which cannot carry non-constant analytic or harmonic functions in the small," *ibid.*, vol. 75 (1953), pp. 260-276.

[3] L. Lichtenstein, "Ueber einige Integrabilitätsbedingungen zweigliedriger Differentialausdrücke mit einer Anwendung auf den Cauchyschen Integralsatz," *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, vol. 9 (1910), pp. 84-100.

[4] ———, "Zur Theorie der konformen Abbildung. Konforme Abbildung nicht-analytischer singularitätenfreier Flächenstücke auf ebene Gebiete," *Bulletin de l'Académie des Sciences de Cracovie*, ser. A (1916), pp. 192-217.

[5] H. Weyl, "Ueber die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement," *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 61 (1915), pp. 40-72.

[6] A. Wintner, "On the Hölder restrictions in the theory of partial differential equations," *American Journal of Mathematics*, vol. 72 (1950), pp. 731-738.

[7] ———, "On Riemann metrics of constant curvature," *ibid.*, vol. 73 (1951), pp. 569-575.

[8] ———, "On isometric surfaces," *ibid.*, vol. 74 (1952), pp. 198-214.

[9] S. Zaremba, "Contribution à la théorie d'une équation fonctionelle de la physique," *Rendiconti del Circolo Matematico di Palermo*, vol. 19 (1905), pp. 140-150.

CORRIGENDA.

Omit the factor ρ^{-1} in all three integrands on p. 273 of [2].

THE GEOMETRY OF THE LINEAR PARTIAL DIFFERENTIAL EQUATION OF THE SECOND ORDER.*

By RICHARD L. INGRAHAM.

1. Introduction. The general linear partial differential equation of the second order in one unknown and n variables has an intrinsic geometry defined by its coefficients. This was investigated first by E. Cotton [1], to whom all the basic results are due. Further work introduced little that was new. Struik and Wiener [2], who were mainly interested in a certain physical application of the Cotton theory, recognized that the geometry of the quadratic and linear differential forms involved, under the groups allowed, could be unified in the concept of one geometry—the Weyl geometry—but otherwise added nothing new to the mathematical theory. Levi-Cività [3], using the Cotton theory, confined himself to the problem of finding normal forms, eliminating one of the most interesting groups by a normalization. Moreover, he paid most attention to the case $n = 2$, an exceptional case to which the general theory does not apply.

The present paper aims first, by making consistent use of the intrinsic Weyl geometry, to cast the known theory in the form in which the powerful transformation calculus of modern differential geometry can be most directly applied to the equivalence problem (which yields a classification) and to the problems of simplifying the equation in the large by suitable transformations and of finding solutions. Second, making use of these methods, it gives several new results, of which the most important is the criterion for the equivalence of two such equations expressed in finite form in terms of complete sets of invariants of the corresponding Weyl geometries. As a corollary, the criterion that an equation be reducible to ordinary Laplacian form is immediate.

2. The intrinsic geometry. We treat only equations with vanishing undifferentiated term. Let the equation be written (cf. [4] for a concise summary of the notations used by Schouten and others)

$$(2.1)\quad g^{rs}\partial^2{}_{rs}\phi + d^r\partial_r\phi = 0 \quad [\partial_r = \partial/\partial x^r;\ r, s, \cdots, = 1, \cdots, n;\ \det g^{rs} \neq 0].$$

* Received March 31, 1953.

The question of whether the left member might represent simply the Laplacian of ϕ in a curved space endowed with a suitable linear connection and metric is answered in the affirmative by the following theorem.

THEOREM 1. *Equations of the type* (2.1) *can always be written as the generalized Laplacian of* ϕ *equals zero in terms of covariant differentiation with respect to a unique Weyl-type linear connection. The associated intrinsic geometry of the quation is a Weyl geometry* W_n $(n \neq 2)$ *and is uniquely determined.*

Proof. The first part of the theorem states that (2.1) is identical with

$$\nabla_r(g^{rs}\partial_s\phi) = 0 \tag{2.2}$$

for a unique Weyl-type connection, i. e., a symmetric linear connection $\Lambda_{pq}{}^t = \Lambda_{qp}{}^t$ for which there exist a symmetric tensor $G_{rs}(\det G_{rs} \neq 0)$ and a vector F_r such that

$$\Lambda_{pq}{}^t = C_{pq}{}^t - \tfrac{1}{2}(A^t{}_pF_q + A^t{}_qF_p - G_{pq}F^t) \tag{2.3}$$

where $C_{pq}{}^t$ is the Christoffel symbol of G_{rs}, G^{rs} are the normalized cofactors, the unit tensor $A^p{}_q \equiv \delta^p{}_q$, and $F^t = G^{tr}F_r$. Expanding (2.2) using (2.3), and comparing with (2.1), one obtains the unique solutions $G_{rs} = g_{rs}$, the normalized cofactors of g^{rs}, and

$$F^r = -2/n(d^r + g^{pq}C_{pq}{}^r), \tag{2.4}$$

which proves the first part. For reasons which will emerge in a moment we define $(n \neq 2)$:

$$f^r \equiv (1 - 2/n)^{-1}F^r = (1 - n/2)^{-1}(d^r + g^{pq}C_{pq}{}^r), \qquad f_s = g_{sr}f^r. \tag{2.4$'$}$$

Then the connection in (2.2) in terms of g_{rs} and f_s is

$$\Lambda_{pq}{}^t = C_{pq}{}^t - \tfrac{1}{2}(1 - 2/n)(A^t{}_pf_q + A^t{}_qf_p - g_{pq}f^t). \tag{2.3$'$}$$

The second part of the theorem asserts that we can really associate a *geometry* with the equation; that is, that the geometrical form in which we have cast (2.1) persists unchanged throughout the group of allowable transformations which take (2.1) into equivalent forms. Now this group [1] is the direct product of $\mathfrak{G}_n$: $x^{p'} = f^{p'}(x^q)$, $\det \partial_q x^{p'} \neq 0$: non-singular change of coordinates; $\mathfrak{F}$: multiplication through of the equation by a factor $\tau(x^q) > 0$:

[1] Beside the two groups considered here, Cotton treats a third group $\phi \to \rho\phi$ transforming the unknown. If this further group is adjoined, the equivalence classes will be correspondingly larger.

gauge group. This second part then asserts that the intrinsic geometry is the Weyl geometry W_n (cf. [5], p. 81) defined by the symmetric and linear differential forms g_{rs} and f_s respectively in the precise sense that a) under a transformation of $\mathfrak{G}_n$, (2.1) goes into $\nabla_{r'}(g^{r's'}\partial_{s'}\phi) = 0$, where the connection $\Lambda_{p'q'}{}^{t'}$ is given by (2.3)′ with

$$\partial_{r'} = \partial/\partial x^{r'}; \qquad g_{r's'} = g_{rs}\partial_{r'}x^r\partial_{s'}x^s, \qquad f_{s'} = f_s\partial_{s'}x^s,$$

and that b) under a transformation of $\mathfrak{F}$ with τ it goes into ${}'\nabla_r({}'g^{rs}\partial_s\phi) = 0$, where the connection ${}'\Lambda_{pq}{}^t$ is given by (2.3)′ with

$${}'g_{rs} = \lambda g_{rs}, \qquad {}'f_s = f_s + \partial_s \log \lambda; \qquad \lambda = \tau^{-1}.$$

For then g_{rs} and f_s transform under $\mathfrak{G}_n \times \mathfrak{F}$ as in W_n.

Proof. a) is immediate from the invariant form of (2.2). b) is shown as follows: $g^{rs} \to {}'g^{rs} \equiv \tau g^{rs}$, hence $g_{rs} \to {}'g_{rs} = \lambda g_{rs}$ for $\lambda = \tau^{-1}$. By (2.4)′,

$$f_s \to {}'f_s \equiv (1 - n/2)^{-1}({}'g_{sr}{}'d^r + {}'g^{pq'}C_{pq}{}^{r'}g_{sr}), \qquad {}'d^r \equiv \tau d^r = \lambda^{-1}d^r.$$

This equals

$$f_s + (2 - n)^{-1}\lambda^{-1}g^{pq}(2g_{ps}\partial_q\lambda - g_{pq}\partial_s\lambda) = f_s + \partial_s \log \lambda,$$

q. e. d. (It should be noted that the connection (2.3)′ in terms of which (2.2) is written is *not* the same as the Weyl connection belonging to W_n. In particular, $\Lambda_{pq}{}^r$ is not gauge-invariant.)

It is remarkable that this theory breaks down for $n = 2$. This case is treated briefly at the end of the article.

Of particular interest is the subclass of self-adjoint equations, those which, after multiplication by a suitable (positive) factor, can be written in the form

(2.1)′ $\qquad \partial_r(P^{rs}\partial_s\phi) = 0 \qquad$ [for some set of functions P^{rs}].

This property has a very nice geometrical characterization (remark: this geometrical characterization of self-adjointness is independent of the boundary behavior of ϕ):

THEOREM 2. Equation (2.1) is self-adjoint if and only if its intrinsic geometry is Riemannian, $(n \neq 2)$.

We recall that W_n is Riemannian if and only if $\partial_{[s}f_{r]} = 0$; i. e., there exists a gauge in which $f_r = 0$.

Proof. If W_n is Riemannian, then in the gauge in which $f_r = 0$ by (2.3)′ ∇_r becomes covariant differentiation with respect to $C_{pq}{}^t$. Hence (2.2) reads

$$\nabla_r(g^{rs}\partial_s\phi) \equiv |\, g\,|^{-\frac{1}{2}}\partial_r(|\, g\,|^{\frac{1}{2}}g^{rs}\partial_s\phi) = 0, \qquad [g \equiv \det g_{rs}].$$

Multiply through by $|\, g\,|^{\frac{1}{2}}$; this proves the sufficiency.

Conversely, if equation (2.1) is self-adjoint, then in a suitable gauge $d^r = \partial_s g^{sr}$. From (2.4)′,

$$f_r = (1 - n/2)^{-1}g_{rs}(\partial_p g^{ps} - \partial_p g^{ps} - \tfrac{1}{2}g^{sq}\partial_q \log|\, g\,|) = (n-2)^{-1}\partial_r \log|\, g\,|,$$

so f_r is a gradient, which proves the necessity.

3. Curvature and the equivalence criterion. Two equations (2.1) are *equivalent* (by this we shall always mean equivalent under the product group $\mathfrak{G}_n \times \mathfrak{F}$) if and only if their intrinsic Weyl geometries are the same. For they both can be written in the form (2.2) with connections of the form (2.3)′ and this prescription is invariant against the transformations considered. It follows that the equivalence problem for these linear equations reduces to the equivalence problem for the Weyl geometry W_n. This closely parallels the usual treatment of the equivalence of quadratic differential forms with slight complications due to the fact that here we have a linear form adjoined and the gauge group as well as coordinate transformation group. It is a noteworthy fact that the equivalence-characterizing system of invariants of the quadratic form and that of the linear form are completely unified in the invariants of W_n. We sketch the proof below, followed by the theorem.

Let g_{pq} and f_r, functions of x^p, define the intrinsic geometry of the first equation, and ${}'g_{p'q'}$, ${}'f_{r'}$, functions of $x^{p'}$, that of the second equation. Then we ask whether there exist a gauge transformation $\lambda(x^{q'})$ and coordinate transformation $x^p = f^p(x^{q'})$ such that

$$(3.1)\qquad {}'g_{p'q'} - \lambda g_{pq}A^p{}_{p'}A^q{}_{q'} = 0, \qquad (3.2)\qquad {}'f_{p'} - f_pA^p{}_{p'} - \Lambda_{p'} = 0,$$

where

$$(3.3)\qquad \partial_{p'}x^p = A^p{}_{p'} \qquad (3.3)'\qquad \partial_{p'}\log\lambda = \Lambda_{p'}.$$

Differentiating (3.1) and using (3.2), (3.3), and (3.3)′, we get on rearranging

$$(3.4)\qquad \Gamma_{p'q'}{}^{r'}A^r{}_{r'} - \Gamma_{pq}{}^rA^p{}_{p'}A^q{}_{q'} - \partial_{p'}A^r{}_{q'} = 0,$$

where $\Gamma_{pq}{}^{r}$ (the linear connection of W_n) is short for

$$\Gamma_{pq}{}^{r} = C_{pq}{}^{r} - \tfrac{1}{2}(A^{r}{}_{p}f_{q} + A^{r}{}_{q}f_{p} - g_{pq}f^{r})$$

and $\Gamma_{p'q'}{}^{r'}$ is the corresponding expression [2] in $'g_{p'q'}$ and $'f_{r'}$. Differentiating (3.2) and using (3.3) and 3.4), we get on rearranging

$$\nabla_{q'}{}'f_{p'} - \nabla_{q}f_{p}A^{q}{}_{q'}A^{p}{}_{p'} - \nabla_{q'}\Lambda_{p'} = 0, \tag{3.4)'}$$

where ∇_q from here on will mean covariant differentiation with respect to $\Gamma_{pq}{}^{r}$, and $\nabla_{q'}$, correspondingly, with respect to $\Gamma_{p'q'}{}^{r'}$. The problem then reduces to solving the system of partial differential equations (3.3), (3.3)′, (3.4), (3.4)′ and the finite equations (3.1), (3.2) in the $(n+1)^2$ unknowns x^{p}, λ, $A^{p}{}_{p'}$, $\Lambda_{p'}$ as functions of $x^{p'}$.

The integrability conditions of (3.3) are satisfied in virtue of (3.4). The integrability conditions of (3.4) are

$$R_{p'q'r'}{}^{t'}A^{t}{}_{t'} = R_{pqr}{}^{t}A^{p}{}_{p'}A^{q}{}_{q'}A^{r}{}_{r'}, \tag{3.5}$$

where $R_{pqr}{}^{t}$ (the curvature tensor of W_n) stands for

$$R_{pqr}{}^{t} \equiv -2\partial_{[p}\Gamma_{q]r}{}^{t} - 2\Gamma_{[p|s|}{}^{t}\Gamma_{q]r}{}^{s}$$

correspondingly for $R_{p'q'r'}{}^{t'}$ in terms of $\Gamma_{q'r'}{}^{t'}$, and the infinite sequence of equations obtained by repeated differentiation of (3.5):

$$\nabla_{s'}R_{p'q'r'}{}^{t'}A^{t}{}_{t'} = \nabla_{s}R_{pqr}{}^{t}A^{p}{}_{p'}A^{q}{}_{q'}A^{r}{}_{r'}A^{s}{}_{s'}, \cdots = \cdots, \text{ etc.} \tag{3.6}$$

It is remarkable to note now that the integrability conditions of the other equations, those arising from the linear form, impose no new conditions. For, the integrability conditions of (3.3)′ are, from (3.4)′,

$$f_{q'p'} = f_{qp}A^{q}{}_{q'}A^{p}{}_{p'}, \qquad f_{qp} \equiv \nabla_{[q}f_{p]} = \partial_{[q}f_{p]},$$

correspondingly for $f_{q'p'}$ in terms of $'f_{p'}$, and the infinite sequence of equations arising from these by repeated differentiation. But multiplying (3.5) through by $A^{s'}{}_{t}$ (the normalized cofactors of $A^{t}{}_{t'}$), contracting r' and s', and using the identities $R_{pqr}{}^{r} = nf_{pq}$, $\nabla_{s}R_{pqr}{}^{r} = n\nabla_{s}f_{pq}, \cdots$, we find that these integrability conditions are satisfied in virtue of (3.5), (3.6), $\cdots$. Moreover, the integrability conditions of (3.4)′, with the aid of (3.2), come out to be

$$(R_{p'q'r'}{}^{t'}A^{t}{}_{t'} - R_{pqr}{}^{t}A^{p}{}_{p'}A^{q}{}_{q'}A^{r}{}_{r'})f_{t} = 0$$

and equations arising from these by repeated differentiation. But these are

[2] Note on notation: Γ instead of $'\Gamma$ is written because it is gauge-invariant (cf. the definition equation). The same remark applies to Δ_q, $R_{pqr}{}^{t}$, and f_{pq}.

satisfied in virtue of (3.5), (3.6), $\cdots$. Hence by the well-known theorem (cf. say, [6], Chap. 5, § 7) on systems of partial differential equations the equivalence theorem for the linear equations (2.1) reads as follows:

THEOREM 3. *Two equations* (2.1) *whose intrinsic geometries are characterized by the invariants* g_{pq}, f_r (*functions of* x^p) *and* $'g_{p'q'}$, $'f_{r'}$ (*functions of* $x^{p'}$) *respectively for* $n \neq 2$ *are equivalent if and only if there exists a positive integer* N *such that* a) *the sets of equations* (3.1), (3.2), *and the first* N *sets of equations* (3.5), (3.6), $\cdots$, *in the unknowns* x^p, λ, $A^p{}_{p'}$, $\Lambda_{p'}$ *as functions of* $x^{p'}$ *are compatible and* b) *all sets of solutions of these equations satisfy the* $(N+1)$-th *set of equations.*

Therewith, equations (2.1) are classified into equivalence classes.

The simplest equation (2.1) is the ordinary Laplacian equation

$$\delta^{rs}\partial^2_{rs}\phi = 0 \qquad [\delta^{rr} = \pm 1; \delta^{rs} = 0, r \neq s]. \tag{3.7}$$

(The term "Laplacian" here embraces all metric signatures.) Then the equivalence theorem gives us immediately the criterion that any equation (2.1) be reducible to this form:

COROLLARY. *An equation* (2.1) *is equivalent to the ordinary Laplacian equation if and only if its intrinsic geometry is flat.*

By *flat* is meant $R_{pqr}{}^t = 0$. (This implies both that the geometry is Riemannian and flat in the Riemannian sense.)

Another application of the complete set of invariants of the intrinsic geometry (in the Riemannian case) is the determination by algebraic means of whether the equation admits any "plane-wave" type solutions. Consider the sets of equations in the unknown $\xi_s(x^p)$

$$g^{rs}\xi_r\xi_s = 0, \tag{3.8}$$

$$R_{pqt}{}^s\xi_s = 0, \tag{3.9}$$

$$\nabla_l R_{pqt}{}^s\xi_s = 0, \tag{3.10}$$

$$\nabla^2{}_{ml} R_{pqt}{}^s\xi_s = 0, \cdots. \tag{3.11}$$

THEOREM 4. *In the Riemannian case* $f_{pq} = 0$, *there exist solutions of* (2.1) *of the plane-wave type if and only if there exists a positive integer* M *such that* a) *the first* M *of* (3.9), (3.10), $\cdots$ *are compatible for the unknown* $\xi_s(x^p)$ *and* b) *all solutions of these satisfy the* $(M+1)$-th *set of equations, finally,* c) *some solution of these satisfies* (3.8).

$\phi = F(\int k_s dx^s)$ is defined to be a *plane-wave* if F is any twice differentiable function, $\int$ means the indefinite line integral, and k_s is a null parallel field: $\nabla_r k_s = 0$, $g^{rs}k_r k_s = 0$.

Proof. If and only if such an M exists does there exist a field k_s parallel with respect to the Riemannian space defined by W_n in the gauge for which $f_r = 0$ (cf. [7], § 23); k_s then is a solution of (3. 9), (3. 10), $\cdots$. If and only if one of these solutions also satisfies (3. 8) does there exist a null parallel field. Hence the conditions of the theorem are necessary and sufficient in order that there exist a plane-wave. But every plane-wave yields a solution of (2. 1). For in the gauge in which $f_r = 0$ we get from (2. 2)

$$g^{rs}\nabla^2{}_{rs}\{F(\int k_s dx^s)\} = F'' g^{rs}k_r k_s = 0,$$

where F' means the derivative of F with respect to its argument.

4. The case $n = 2$. We add a few words on the anomalous case $n = 2$. The first part of Theorem 1 is still true, but no Weyl geometry W_2 can be associated with the equation. For although g_{pq} and F_r (given by (2. 4)) are tensors against $\mathfrak{G}_2$, under $\mathfrak{F}$ they transform as follows:

$$'g_{pq} = \lambda g_{pq}, \qquad 'F_p = F_p. \tag{4. 1}$$

Hence the geometry is that of a class of conformally related Riemannian spaces V_2, on each of which the same Pfaffian is superimposed.

Theorem 2 holds in the form: Equation (2. 1) is self-adjoint if and only if F_r is a gradient.

Of course $\partial_{[s}F_{r]} = 0$ has not now any Riemannian interpretation. In proof, note that if the equation is self-adjoint, then in a coordinate and gauge frame in which it takes the form (2. 1)', we have $F_r = \frac{1}{2}\partial_r \log |g|$ (cf. the proof of Theorem 2). Conversely, if F_r is a gradient, $\frac{1}{2}\partial_r \log h$, $h > 0$, say, perform the gauge transformation with $\lambda = h^{\frac{1}{2}} |g|^{-\frac{1}{2}}$ to a frame where $|'g| = h$, $'F_r = F_r = \frac{1}{2}\partial_r \log |'g|$. Then, as in the proof of Theorem 2, in this coordinate and gauge frame $'d^r = \partial_s {}'g^{sr}$ and hence the equation is self-adjoint.

Theorem 3 does not apply. In the present case $n = 2$ the equations we start from in the equivalence problem are

$$'g_{p'q'} - \lambda g_{pq}A^p{}_{p'}A^q{}_{q'} = 0; \qquad \partial_{p'}x^p = A^p{}_{p'} \tag{4. 2}$$

$$'F_{p'} - F_p A^p{}_{p'} = 0. \tag{4. 3}$$

Since any two V_2's of the same signature are conformal (cf. [7], § 28), the problem reduces to using the remaining freedom in the coordinate transformations satisfying (4. 2) to satisfy (4. 3) as well. This was treated at lenth by Cotton ([1], p. 236 *et seq.*) and also elsewhere in the literature.

From the facts that $F_r = 0$ for the ordinary Laplacian equation (3.7), F_r is gauge-invariant for $n = 2$, and any two V_2's of the same signature are conformal, we infer immediately the following

COROLLARY. *The vanishing of F_r is necessary and sufficient in order that the equation* (2.1) *be reducible to ordinary Laplacian form.*

Thus in the case of a definite metric, $F_r = 0$ is the criterion that the solution of (2.1) be reducible to the solution of the Beltrami differential equations.

INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

[1] E. Cotton, "Sur les invariants différentiels de quelques équations linéaires aux dérivées partielles du second ordre," *Annales de l'Ecole Normale Supérieure*, vol. 17 (1900), pp. 211-244.

[2] D. Struik and N. Weiner, "A relativistic theory of quanta," *Journal of Mathematics and Physics*, vol. 7 (1927), pp. 1-23.

[3] T. Levi-Cività, "Sulla trasformazione delle equazioni lineari a derivate parziali del secondo ordine," *Atti del Reale Istituto Veneto*, 3a ser. 15 tomo 72 (1912-13), pp. 1331-1357.

[4] J. Shouten, "La théorie projective de la relativité," *Annales de L'Institut Henri Poincaré*, vol. 5 fasc. 1 (1935), pp. 51-88.

[5] L. Eisenhart, "*Non-Riemannian Geometry*," (1927).

[6] O. Veblen, "*Invariants of Quadratic Differential Forms*," (1927).

[7] L. Eisenhart, "*Riemannian Geometry*," (1926).

KUMMER CONGRUENCES AND THE SCHUR DERIVATIVE.*

By L. CARLITZ.

1. Introduction. Let p be a prime, a an integer not divisible by p, and put

$$\Delta a^{p^m} = (a^{p^{m+1}} - a^{p^m})/p^{m+1}, \qquad \Delta^r a^{p^m} = \Delta(\Delta^{r-1} a^{p^m}) \quad (r = 2, 3, \cdots). \tag{1.1}$$

Schur [6] proved that the derivatives $\Delta a^{p^m}, \Delta^2 a^{p^m}, \cdots, \Delta^{p-1} a^{p^m}$ are all integral. If $a^{p-1} \equiv 1 \pmod{p^2}$ then all the derivatives $\Delta^r a^{p^m}$ are integral, while if $a^{p-1} \not\equiv 1 \pmod{p^2}$, then every number $\Delta^p a^{p^m}$ has exactly the denominator p. A. Brauer [1] gave another proof of these results; about the same time Zorn [7] also proved these and other results by p-adic methods.

In the present paper we consider sequences of rational numbers $\{a_m\}$ that are integral (mod p), where p is a fixed prime. (More generally the a_m may be integral p-adic numbers.) Now suppose that the a_m satisfy

$$\sum_{s=0}^{r} (-1)^{r-s} C_s^r a_p^{r-s} a_{m+s(p-1)} \equiv 0 \pmod{p^r} \tag{1.2}$$

for $m \geqq r \geqq 1$. We shall call (1.2) Kummer's congruence for $\{a_m\}$. The same term may also be used for the stronger congruence

$$\sum_{s=0}^{r} (-1)^{r-s} C_s^r a_p^{(r-s)b/(p-1)} a_{m+sb} \equiv 0 \pmod{p^{re}}, \tag{1.3}$$

where $p^{e-1}(p-1) \mid b$ and $m \geqq re$, $e \geqq 1$, $r \geqq 1$. As we shall see, (1.2) implies (1.3).

Next, generalizing (1.1), we define

$$\Delta a_{p^m} = (a_{p^{m+1}} - a_p{}^{p^m} a_{p^m})/p^{m+1},$$
$$\Delta^r a_{p^m} = (\Delta^{r-1} a_{p^{m+1}} - a_p{}^{p^{m+r-1}} \Delta^{r-1} a_{p^m})/p^{m+1}. \tag{1.4}$$

Then we shall show that if the sequence $\{a_m\}$ satisfies (1.2) it follows that the numbers

$$\Delta a_{p^m}, \Delta^2 a_{p^m}, \cdots, \Delta^{p-1} a_{p^m} \tag{1.5}$$

are integral for all m. Moreover the residues of these numbers (mod p^m) are specified. The method of proof is essentially that used in [4]. We also

* Received October 16, 1952.

consider the vase in which the a_m are polynomials in an indeterminate u with integral coefficients.

We also discuss some applications (§§ 6, 7) of these results as well as a generalization (§ 8).

2. Kummer's congruences. To show that (1.2) implies (1.3) we remark first of all that by the binomial expansion

$$(2.1)\qquad t^{p(p-1)} - u^p = (t^{p-1} - u)^p + p(t^{p-1} - u)f(t, u),$$

where $f(t, u)$ denotes a polynomial in the indeterminates t, u with integral coefficients. Then by a straightforward induction (2.1) yields the more general formula

$$(2.2)\qquad t^{p^e(p-1)} - u^{p^e} = (t^{p-1} - u)^{p^e} + \sum_{i=1}^{e} p^i (t^{p-1} - u)^{p^{e-i}} f_i(t, u)$$

for all $e \geqq 1$; the $f_i(t, u)$ are polynomials with integral coefficients. We shall also require the formula obtained from (2.2) by raising both members to the r-th power and expanding the right side; this may be referred to as $(2.2)_r$.

Now returning to (1.2) it is clear that the left side may be written in the following form:

$$(2.3)\qquad (E^{p-1} - a_p)^r a_m \qquad (Ea_m = a_{m+1}),$$

provided we agree that E operates only on m (and therefore commutes with a_p). Similarly the left side of (1.3) may be written

$$(2.4)\qquad (E^b - a_p{}^{b/(p-1)})^r a_m.$$

Let us consider first the case $e = 1$, so that $b = c(p-1)$, $c \geqq 1$. Comparing (2.3) with (2.4) it is clear since $E^{c(p-1)} - a_p{}^c$ contains $E^{p-1} - a_p$ as a factor that (1.3) holds in this case. For arbitrary $e \geqq 1$, we take $t = E^{c(p-1)}$, $u = a_p{}^c$ in $(2.2)_r$ and apply the special case of (1.3) just obtained. Each term in the right member is seen to be divisible by at least p^{er}. This proves

THEOREM 1. *If* (1.2) *holds for* $m \geqq r \geqq 1$ *then* (1.3) *also holds for all* $m \geqq er$, $r \geqq 1$, $e \geqq 1$.

For some applications we shall be interested in the following extension. Let $a_m = a_m(u)$ denote a polynomial in an indeterminate u with rational coefficients that are integral (mod p). (If $f(u)$ is such a polynomial, then the statement $f(u) \equiv 0 \pmod{p^r}$ means that each coefficient of $f(u) \equiv 0 \pmod{p^r}$.) It is then clear what meaning is to be attached to (1.2) and

(1.3) in this situation. Moreover it is evident that the above proof applies. We may state

THEOREM 1′. *If the a_m are polynomials in u with integral coefficients, then Theorem 1 holds.*

We also remark that if in (1.2) we replace a_p by a number congruent to it (mod p), then the congruence will continue to hold; the same is therefore true of (1.3). Indeed we have the identity

$$\sum_{s=0}^{r}(-1)^{r-s}C_s{}^r a_{m+s(p-1)}(a_p+t)^{r-s} = \sum_{i=0}^{r}(-1)^i C_i{}^r t^i \sum_{s=0}^{r-i}(-1)^s C_s{}^{r-i} a_{m+s(p-1)} a_p{}^{r-i-s},$$

from which the above statement is evident.

3. Some lemmas. We shall require the following lemmas.

LEMMA 1.

$$\prod_{i=0}^{r-1}(x-p^i y) = \sum_{i=0}^{r}(-1)^i [r,i] p^{\frac{1}{2}i(i-1)} x^{r-i} y^i, \tag{3.1}$$

where $[r,i] = \dfrac{(p^r-1)\cdots(p^{r-i+1}-1)}{(p-1)\cdots(p^i-1)} = [r,r-1],\ [r,0]=1.$

LEMMA 2. *If* $e_i = (p^i-1)/(p-1)$, $e_0 = 0$, *then*

$$\Delta^r a_{p^m} = p^{-rm-\frac{1}{2}r(r+1)} \sum_{i=0}^{r}(-1)^i [r,i] p^{\frac{1}{2}i(i-1)} a_{p^{m+r-i}} a_p{}^{p^{m+r-i}e_i}. \tag{3.2}$$

LEMMA 3. *Let*

$$W_{k,r} = \sum_{i=0}^{k}(-1)^{k-i}[k,i] C_r{}^{e_i} p^{\frac{1}{2}(k-i)(k-i-1)},$$

where $e_i = (p^i-1)/(p-1)$ *and* $C_r{}^m$ *denotes a binomial coefficient. Then*

$$W_{k,r} = 0,\ p^{\frac{1}{2}r(r-1)} \prod_{i=1}^{r} e_i/r!,\ p^{\frac{1}{2}k(k-1)} U_{k,r}/r! \tag{3.3}$$

according as $r<k$, $r=k$, $r>k$, *respectively, where* $U_{k,r}$ *is an integer.*

Lemma 1 is familiar. Lemma 2 is a slight extension of a formula of Schur. For Lemma 3, compare [4], Lemma 2.

4. A formula for $\Delta^r a_{p^m}$. If in (1.3) we take $b = p^e(p-1)$, $r=1$, $m=p^e$, we get $a_{p^{e+1}} - a_p{}^{p^e} a_{p^e} \equiv 0 \pmod{p^{e+1}}$. In other words, using the

first of (1.4), we see that Δa_{p^e} is integral. In order to treat the general case we replace b and m in (1.3) by $p^m(p-1)$ and p^m, respectively, and put

$$\sum_{s=0}^{r}(-1)^{r-s}C_s^r a_p^{(r-s)p^m} a_{p^m(1+s(p-1))} = p^{r(m+1)}Q_{m,r}, \tag{4.1}$$

so that $Q_{m,r}$ is integral (mod p). Now it is easily verified that (4.1) implies

$$a_{p^m(1+r(p-1))} = \sum_{s=0}^{r} C_s^r a_p^{(r-s)p^m} p^{s(m+1)} Q_{m,s}. \tag{4.2}$$

In particular for $r = e_k = (p^k-1)/(p-1)$, (4.2) becomes

$$a_{p^{m+k}} = \sum_{s=0}^{e_k} C_s^{e_k} a_p^{(e_k-s)p^m} p^{s(m+1)} Q_{m,s}. \tag{4.3}$$

Substituting from (4.3) in (3.2) we get

$$\begin{aligned} p^{rm+\frac{1}{2}r(r+1)}\Delta^r a_{p^m} &= \sum_{i=0}^{r}(-1)^{r-i}p^{\frac{1}{2}(r-i)(r-i-1)}a_p^{p^{m+i}}e_{r-i}[r,i] \\ &\qquad \cdot \sum_{s=0}^{e_i} C_s^{e_i} a_p^{(e_i-s)p^m} p^{s(m+1)} Q_{m,s} \\ &= \sum_{s=0}^{e_r} p^{s(m+1)} a_p^{(e_r-s)p^m} Q_{m,s} \sum_{i=0}^{r}(-1)^{r-i}[r,i]\, C_s^{e_i} p^{\frac{1}{2}(r-i)(r-i-1)} \\ &= \sum_{s=0}^{e_r} p^{s(m+1)} a_p^{(e_r-s)p^m} Q_{m,s} W_{r,s} \\ &= (1/r!)\, p^{r(m+1)+\frac{1}{2}r(r-1)} a_p^{(e_r-r)p^m} Q_{m,r} \prod_{i=1}^{r} e_i \\ &\quad + \sum_{s=r+1}^{e_r} (1/s!)\, p^{s(m+1)+\frac{1}{2}r(r-1)} a_p^{(e_r-s)p^m} Q_{m,s} U_{m,s}, \end{aligned}$$

by (3.3); $W_{m,s}$ and $U_{m,s}$ have the same meaning as in Lemma 3. Thus it follows that

$$\begin{aligned} \Delta^r a_{p^m} &= (1/r!)\, a_p^{(e_r-r)p^m} Q_{m,r} \prod_{i=1}^{r} e_i \\ &\quad + \sum_{s=r+1}^{e_r} (1/s!)\, p^{(m+1)(s-r)} a_p^{(e_r-s)p^m} Q_{m,s} U_{m,s}. \end{aligned} \tag{4.4}$$

It may be of interest to mention a variant of (4.4). In (1.3) we replace both b and m by $p^m(p-1)$ and put

$$\sum_{s=0}^{r}(-1)^{r-s}C_s^r a_p^{(r-s)p^m} a_{p^m(s+1)(p-1)} = p^{r(m+1)}Q'_{m,r}, \tag{4.5}$$

so that $Q'_{m,r}$ is integral (mod p). Then (4. 5) implies

$$a_{p^m(r+1)(p-1)} = \sum_{s=0}^{r} C_s{}^r a_p{}^{(r-s)p^m} p^{s(m+1)} Q'_{m,s}. \tag{4. 6}$$

In particular for $r+1 = p^k$, (4. 6) becomes

$$a_{(p-1)p^{m+k}} = \sum_{s=0}^{p^k-1} C_s{}^{p^k-1} a_p{}^{(p^k-1-s)p^m} p^{s(m+1)} Q'_{m,s}. \tag{4. 7}$$

Substituting from (4. 7) in (1. 3) we get, very much as before,

$$\Delta^r a_{(p-1)p^m} = (1/r!) a_p{}^{(p^r-1-r)p^m} Q'_{m,r} \prod_{i=1}^{r} (p^i - 1) + \sum_{s=r+1}^{p^r-1} (1/s!) p^{(m+1)(s-r)} a_p{}^{(p^r-1-s)p^m} Q'_{m,s} U'_{m,s}, \tag{4. 8}$$

where $U'_{m,s}$ is integral and

$$\Delta a_{(p-1)p^m} = (a_{(p-1)p^{m+1}} - a_p{}^{(p-1)p^m} a_{(p-1)p^m})/p^{m+1},$$

and

$$\Delta^r a_{(p-1)p^m} = (\Delta^{r-1} a_{(p-1)p^{m+1}} - a_p{}^{(p-1)p^{m+r-1}} \Delta^{r-1} a_{(p-1)p^m})/p^{m+1}.$$

Additional formulas of this kind are easily obtained.

5. The main results. By means of (4. 4) it is easy to derive the main results concerning $\Delta^r a_{p^m}$ (compare [4], § 3). It is evidently only necessary to examine $p^{(m+1)(s-r)}/s!$ $(s > r)$. Let $r \leqq p$; then $p^{s-r}/s!$ is integral (mod p) and is indeed divisible by p unless (i) $s = p$, $r = p-1$, or (ii) $s = p+1$, $r = p$. We may now state the following two theorems.

THEOREM 2. *$\Delta^r a_{p^m}$ is integral for $1 \leqq r \leqq p-1$, while $\Delta^p a_{p^m}$ has the denominator p provided $Q_{m,p} \not\equiv 0$ (mod p).*

THEOREM 3. For $1 \leqq r \leqq p$,

$$\Delta^r a_{p^m} \equiv a_p{}^{(e_r-r)p^m} \frac{Q_{m,r} \prod_{i=1}^{r} (p^i - 1)}{r!(p-1)^r} \pmod{p^m}, \tag{5. 1}$$

where $Q_{m,r}$ is defined by (4. 1).

Similarly using (4. 8) we get, for $1 \leqq r \leqq p$,

$$\Delta^r a_{(p-1)p^m} \equiv (1/r!) a_p{}^{(p^r-1-r)p^m} Q'_{m,r} \prod_{i=1}^{r} (p^i - 1) \pmod{p^m}, \tag{5. 2}$$

where $Q'_{m,r}$ is defined by (4. 5); for $r < p-1$, (5. 2) holds (mod p^{m+1}).

In the next place, if as in Theorem 1′ we suppose the a_m polynomials

in u with integral coefficients then Theorems 2 and 3 will continue to hold. The corresponding results will be referred to as Theorem 2′ and Theorem 3′. It seems unnecessary to state these theorems explicitly.

6. Applications to the coefficients of the Jacobi ellipic functions. We now assume $p > 2$. The writer has proved [2] that if $l = k^2$ is rational (mod p), and

$$sn\, x = sn(x, k^2) = \sum_{m=1}^{\infty} a_m x^m/m!, \tag{6.1}$$

then the coefficients a_m satisfy (1.2). Consequently we have at once

THEOREM 4. *The coefficients a_m defined by* (6.1) *satisfy Theorems* 2 *and* 3.

However it is evident that we can say a great deal more. For not only does a_m satisfy (1.3) but so also does a_{m+t} and a_{mt}. Accordingly we state

THEOREM 5. *$\Delta^r a_{p^m t}$ and $\Delta^r a_{p^m+t}$ are integral* (mod p), *where t is an arbitrary integer.*

It is also easily seen that the last two theorems hold for the coefficients b_m defined by

$$sn^h x\, cn^{h'} x\, dn^{h''} x = \sum_{m=h}^{\infty} b_m x^m/m!, \tag{6.2}$$

where h, h', h'' are integers and $h \geqq 0$.

As for the coefficients β_m defined by

$$x/sn\, x = \sum_{m=0}^{\infty} \beta_m x^m/m! \tag{6.3}$$

we have by [2], Theorem 4

$$\sum_{i=0}^{r} (-1)^i C_i^r a_p^{r-i} \tau_{m+i(p-1)} \equiv 0 \pmod{p^r}, \tag{6.4}$$

where $\tau_m = \beta_m/m$, $p - 1 \nmid m$ and $m > r \geqq 1$. Comparison of (6.4) with (1.2) shows that the latter is not quite satisfied in this case. However it is easily verified that the following theorem holds.

THEOREM 6. *Let $\tau_m = \beta_m/m$, where β_m is defined by* (6.3). *Then*

$$\Delta^r \tau_{k+p^m} \qquad (r < p, r \leqq m, k \geqq 1) \tag{6.5}$$

is integral provided $p - 1 \nmid k + 1$.

In the next place if in (6.1) we use the fuller notation

$$sn(x, u) = \sum_{m=1}^{\infty} A_m(u)x^m/m!, \tag{6.6}$$

where $A_m(u)$ is a polynomial in the indeterminate u with integral coefficients. Then by [3], Theorem 3, we see that $A_m(u)$ also satisfies the congruence (1.2). Thus Theorem 1′ as well as the later discussion applies and we obtain

THEOREM 7. *The coefficients $A_m(u)$ defined by* (6.4) *satisfy Theorems 2′ and 3′.*

It is clear how the results concerning (6.2) and (6.5) can be carried over to the present situation; we shall accordingly not take the space to state these theorems explicitly.

7. Application to Eulerian polynomials. For a second application we consider the Eulerian polynomials $A_k(u)$ which may be defined by means of $A_k(u) = (u-1)^k H_k(u)$, where

$$(1-u)/(e^x - u) = \sum_{k=0}^{\infty} H_k(u)x^k/k!.$$

Then Frobenius [5] has proved that $H_k(u)$ satisfies a congruence of the form (1.3), from which it follows that the polynomial $A_k(u)$ also satisfies a congruence of the same sort. Consequently applying Theorem 2′ we can assert that the polynomials

$$\Delta^r A_{kp^m}(u), \qquad \Delta^r A_{k+p^m}(u) \tag{7.1}$$

have integral coefficients for $r < p$. We can also specify the residues of (7.1) (mod p^m) in terms of the corresponding Kummer quotients.

8. A generalization. As in [4, § 5] the definition (1.4) can be generalized. Let us define

$$\begin{aligned} \Delta_p a_{mp^i} &= (a_{mp^{i+1}} - a_p{}^{mp^i} a_{mp^i})/p^{i+1}, \\ \Delta_p{}^r a_{mp^i} &= (\Delta_p{}^{r-1} a_{mp^{i+1}} - a_p{}^{mp^{i+r-1}} \Delta_p{}^{r-1} a_{mp^i})/p^{i+1}. \end{aligned} \tag{8.1}$$

Then it is clear that $\Delta_p{}^r \Delta_q{}^s = \Delta_q{}^s \Delta_p{}^r$. Now put

$$\delta_k{}^{(r)} a_{mk} = \Delta_{p_1}{}^{r_1} \cdots \Delta_{p_s}{}^{r_s} a_{mk}, \tag{8.2}$$

where $k = p_1{}^{e_1} \cdots p_s{}^{e_s}$. Theorem 2 implies the following generalization.

THEOREM 8. *Let k be a fixed integer $\geqq 2$. If (1.2) holds for all primes p_i dividing k and $r_i < p_i$, $j = 1, \cdots, s$, then $\delta_k a_k$ is integral.*

In particular when all the r_i are equal we may write $\delta_k{}^r$ in place of $\delta_k{}^{(r)}$ in (8.2); it is possible to set up explicit formulas for both $\delta_k{}^r$ and $\delta_k{}^{(r)}$ (compare [4], § 5).

DUKE UNIVERSITY.

REFERENCES.

[1] A. Brauer, "Über eine Erweiterung des kleinen Fermatschen Satzes," *Mathematische Zeitschrift*, vol. 42 (1937), pp. 255-262.

[2] L. Carlitz, "Congruences for the coefficients of the Jacobi elliptic functions," *Duke Mathematical Journal*, vol. 16 (1949), pp. 297-302.

[3] ———, "Congruences connected with the power series expansions of the Jacobi elliptic functions," *Duke Mathematical Journal*, vol. 20 (1953), pp. 1-12.

[4] ———, "Some theorems on the Schur derivative," *Pacific Journal of Mathematics*, vol. 3 (1953), pp. 321-332.

[5] G. Frobenius, "Über die Bernoullischen Zahlen und die Eulerschen Polynome," *Sitzungsberichte der Preussischen Akademie der Wissenschaften* (1910), pp. 809-847.

[6] I. Schur, "Ein Beitrag zur elementaren Zahlentheorie," *Sitzungsberichte der Preussischen Akademie der Wissenschaften* (1933), pp. 145-151.

[7] M. Zorn, "p-adic analysis and elementary number theory," *Annals of Mathematics* (2), vol. 38 (1937), pp. 451-464.

SOME CONGRUENCES OF VANDIVER.*

By L. CARLITZ.

---

1. Introduction. Professor Vandiver has kindly called the writer's attention to the following congruence for the Bernoulli numbers:

$$h_1^{n_1}\cdots h_k^{n_k}(\lambda_1 h_1^{p-1}+\cdots+\lambda_k h_k^{p-1})^r \equiv 0 \pmod{p^r, p^{n_1-1},\cdots,p^{n_k-1}} \tag{1.1}$$

$$(n_i \not\equiv 0 \pmod{p-1}, i=1,\cdots,k),$$

where the left-member is expanded in full and B_m/m substituted for h_i^m in the result; B_m is the Bernoulli number in the even suffix notation ($(B+1)^m = B_m$, $m>1$), the λ's are rational integers such that

$$\lambda_1+\cdots+\lambda_k \equiv 0 \pmod{p}, \tag{1.2}$$

and p is an odd prime. For example (1.1) implies in particular

$$\sum_{s=0}^{r}(-1)^s C_s^r \frac{B_{m+(r-s)(p-1)}}{m+(r-s)(p-1)}\,\frac{B_{n+s(p-1)}}{n+s(p-1)} \equiv 0 \pmod{p^r}$$

provided $p-1\nmid m$, $p-1\nmid n$, $m>r$, $n>r$. The congruence (1.1) is derived from a more general result proved in [5].

In the present note we wish to point out a generalization of (1.1) that is valid for certain sequences including the Bernoulli and Euler numbers. Let $\{a_m\}$ be a sequence of rational numbers that are integral (mod p) and assume that

$$\sum_{s=0}^{r}(-1)^{r-s}C_s^r a_{m+s(p-1)}a'^{r-s}_p \equiv 0 \pmod{p^r, p^m}, \tag{1.3}$$

where a'_p is integral (mod p). We may call (1.3) Kummer's congruence for $\{a_m\}$. Now let $\{b_m\}$ denote another sequence that satisfies

$$\sum_{s=0}^{r}(-1)^{r-s}C_s^r b_{m+s(p-1)}b'^{r-s}_p \equiv 0 \pmod{p^r, p^m},$$

where b'_p is integral (mod p). Then we prove that

$$\sum_{s=0}^{r} C_s^r a_{m+(r-s)(p-1)}b_{n+s(p-1)}(\mu a'_p)^s(\lambda b'_p)^{r-s} \equiv 0 \pmod{p^r, p^m, p^n}, \tag{1.4}$$

---

* Received January 16, 1953.

provided $\lambda + \mu \equiv 0 \pmod{p}$. The left member of (1.4) can be written briefly as

$$a^m b^n (\lambda a^{p-1} b'_p + \mu b^{p-1} a'_p)^r. \tag{1.5}$$

More generally we prove that if $p^{e-1}(p-1) \mid u$, $p^{e-1}(p-1) \mid v$ and $\lambda_1 + \cdots + \lambda_k \equiv 0 \pmod{p^e}$, then

$$a^m b^n (\lambda a^u b'^{v/(p-1)}_p + \mu b^v a'^{u/(p-1)}_p)^r \equiv 0 \pmod{p^{re}, p^m, p^n}, \tag{1.6}$$

where it is understood that the left member of (1.6) is to be expanded in full and $a^{m+(r-s)u}$, b^{n+sv} replaced by $a_{m+(r-s)u}$, b_{n+sv}, respectively. For the generalization of (1.6) corresponding to (1.1), see (3.5) below.

We also indicate a few applications of these results, particularly to the coefficients of the Jacobi elliptic functions, in § 4.

For some related results on sequences satisfying Kummer's congruences, see [3].

2. Proof of (1.4) and (1.6). To prove (1.4) we require the following identity

$$a^m b^n (\lambda a^{p-1} b'_p + \mu b^{p-1} a'_p)^r = \sum_{i+j \leqq r} \frac{r!}{i!\, j!\, (r-i-j)!} \text{ times} \tag{2.1}$$

$$a^m (a^{p-1} - a'_p)^i b^n (b^{p-1} - b'_p)^j (\lambda b'_p)^i \cdot (\mu a'_p)^j (\lambda + \mu)^{r-i-j} (a'_p b'_p)^{r-i-j}.$$

Indeed (2.1) is almost immediate if we raise both members of the equation

$$\lambda a^{p-1} b'_p + \mu b^{p-1} a'_p = (a^{p-1} - a'_p)\lambda b'_p + \mu (b^{p-1} - b'_p) a'_p + (\lambda + \mu) a'_p b'_p$$

to the r-th power and subsequently multiply by $a^m b^n$. If we prefer, we can give a straightforward proof of (2.1) without the use of symbolic notation.

Now assume that $\{a_m\}$ satisfies (1.3) and that $\{b_m\}$ satisfies a like congruence; also that $\lambda + \mu \equiv 0 \pmod{p}$. Then it is clear that each term in the right member of (2.1) is $\equiv 0 \pmod{p^i \cdot p^j \cdot p^{r-i-j}, p^m, p^n}$. Thus it follows that

$$a^m b^n (\lambda a^{p-1} b'_p + \mu b^{p-1} a'_p)^r \equiv 0 \pmod{p^r, p^m, p^n}, \tag{2.2}$$

so that we have proved (1.4).

In the next place we remark that if the sequence $\{a_m\}$ satisfies (1.3) then indeed the following congruence holds:

$$a^m (a^w - a'^{w/(p-1)}_p)^r \equiv 0 \pmod{p^{re}, p^m}, \tag{2.3}$$

where $p^{e-1}(p-1) \mid w$. (For proof see [2], Theorem 1; while in that theorem

a'_p was taken equal to a_p, this is not required in the proof.) In place of (2.1) we now employ the identity

$$(2.4)\qquad a^m b^n(\lambda a^u {b'_p}^{v/(p-1)} + \mu b^v {a'_p}^{u/(p-1)})^r$$
$$= \sum_{i+j\leqq r} \frac{r!}{i!\,j!\,(r-i-j)!}\, a^m(a^u - {a'_p}^{u/(p-1)})^i b^n(b^v - {b'_p}^{v/(p-1)})^j$$
$$\times (\lambda {b'_p}^{v/(p-1)})^i(\mu {a'_p}^{u/(p-1)})^j(\lambda+\mu)^{r-i-j}({a'_p}^{u/(p-1)}{b'_p}^{v/(p-1)})^{r-i-j}.$$

As above, (2.4) follows readily from

$$\lambda a^u {b'_p}^{v/(p-1)} + \mu b^v {a'_p}^{u/(p-1)} = (a^u - {a'_p}^{u/(p-1)})\lambda {b'_p}^{v/(p-1)}$$
$$+ (b^v - {b'_p}^{v/(p-1)})\mu {a'_p}^{u/(p-1)} + (\lambda+\mu){a'_p}^{u/(p-1)}{b'_p}^{v/(p-1)}.$$

Hence with the assumptions used in proving (2.2) we have first that $\{a_m\}$ satisfies (2.3) and that $\{b_m\}$ satisfies a like congruence. Assuming also that $\lambda+\mu \equiv 0 \pmod{p^e}$ it is clear that (2.4) implies

$$a^m b^n(\lambda a^u {b'_p}^{v/(p-1)} + \mu b^v {a'_p}^{u/(p-1)})^r \equiv 0 \pmod{p^{re}, p^m, p^n},$$

so that we have proved (1.6).

Note that (1.6) continues to hold if we replace a'_p, b'_p by a''_p, b''_p, respectively, where $a'_p \equiv a''_p$, $b'_p \equiv b''_p \pmod p$. This is a fairly easy consequence of the fact that ${a'_p}^{p^{e-1}} \equiv {a''_p}^{p^{e-1}}$, ${b'_p}^{p^{e-1}} \equiv {b''_p}^{p^{e-1}} \pmod{p^e}$, so that the left member of (1.6) is unchanged $\pmod{p^e}$ by the replacement.

3. The general case. The statement of the general case corresponding to (1.1) and (1.6) is somewhat complicated notationally. Let $\{a_{i,m}\}$, $i = 1, \cdots, k$, denote k sequences each satisfying

$$(3.1)\qquad {a_i}^m({a_i}^{p-1} - a'_{i,p})^r \equiv 0 \pmod{p^r, p^m},$$

where after expansion of the left member ${a_i}^{m+s(p-1)}$ is replaced by $a_{i,m+s(p-1)}$. Let $\lambda_1, \cdots, \lambda_k$ denote rational numbers that are integral $\pmod p$ and satisfy

$$(3.2)\qquad \lambda_1, \cdots, \lambda_k \equiv 0 \pmod{p^e}.$$

Let $u_1, \cdots, u_k$ denote integers such that

$$(3.3)\qquad p^{e-1}(p-1) \mid u_i \qquad (i = 1, \cdots, k).$$

Also put

$$(3.4)\qquad c_{i,p} = \prod_{j=1}^{k}{}' {a'_{i,p}}^{u_i/(p-1)}, \text{ where } j \neq i.$$

We now state the following

THEOREM 1. *If* (3.1), (3.2), (3.3) *and* (3.4) *hold, then*

$$a_1^{m_1}\cdots a_k^{m_k}(\lambda_1 a_1^{u_1}c_{1,p}+\cdots+\lambda_k a_k^{u_k}c_{k,p})^r\equiv 0 \pmod{p^{re}, p^{m_1},\cdots,p^{m_k}}, \tag{3.5}$$

where after expansion of the left member $a_i^{m_i+su_i}$ *is replaced by* a_{i,m_i+su_i}.

The proof of (3.5) is very much like the proof of (1.6) except that (2.4) is replaced by a somewhat more elaborate formula, the basis of which is the evident identity

$$\sum_{i=1}^{k}\lambda_i a_i^{u_i}c_{i,p}=\sum_{i=1}^{k}\lambda_i(a_i^{u_i}-a'^{u_i/(p-1)}_{i,p})c_{i,p}+\sum_{i=1}^{k}\lambda_i\prod_{j=1}^{k}a'^{u_i/(p-1)}_{i,p}.$$

In view of the complicated nature of the hypotheses leading to (3.5) as well as of the formula itself, it may be of interest to state explicitly the following special case which indeed corresponds more closely to (1.1). We assume (3.1) but in place of (3.2), (3.3), (3.4) we suppose that

$$\lambda_1+\cdots+\lambda_k\equiv 0 \pmod{p}, \tag{3.2}'$$

$$c_{i,p}=\prod_{j=1}^{k}{}' a'_{i,p}, \text{ where } j\neq i. \tag{3.4}'$$

Then we can assert that if the k sequences $\{a_{i,m}\}$ satisfy (3.1), (3.2)′ and (3.4)′, it follows that

$$a_1^{m_1}\cdots a_k^{m_k}(\lambda_1 a_1^{p-1}c_{1,p}+\cdots+\lambda_k a_k^{p-1}c_{k,p})^r\equiv 0 \pmod{p^r, p^{m_1},\cdots,p^{m_k}}. \tag{3.5}'$$

As a corollary of Theorem 1 we state

THEOREM 2. *Assume in addition to the hypothesis of Theorem* 1 *that* $a'_{1,p}\equiv\cdots\equiv a'_{k,p} \pmod{p}$. *Then we have*

$$a_1^{m_1}\cdots a_k^{m_k}(\lambda_1 a_1^{u_1}+\cdots+\lambda_k a_k^{u_k})^r\equiv 0 \pmod{p^{re}, p^{m_1},\cdots,p^{m_k}}. \tag{3.6}$$

In particular Theorem 2 applies when the sequences $\{a_{i,m}\}$ are identical.

To prove the theorem we note first that (3.6) is an immediate consequence of (3.5) when $p\nmid a'_p$. On the other hand when $p\mid a'_p$, it follows readily from (2.3) and the identity

$$a^{m+sw}=\sum_{t=0}^{s}C_t^s a^m(a^w-a'^{w/(p-1)}_p)^{s-t}a'^{tw/(p-1)}_p$$

that $a_{m+sw}\equiv 0 \pmod{p^{se}}$ and this in turn leads to (3.6) in the case $p\mid a'_p$.

4. Some applications. In the first place it is clear that (1.6) and (3.5) hold for the Euler numbers E_m (in the even suffix notation) and for the numbers $h_m = B_m/m$ occurring in (1.1) as well as for various related sequences. Moreover these numbers can be combined in various ways. For example we may mention the special case

$$h^m E^n (h^{p-1} - E^{p-1})^r \equiv 0 \pmod{p^r, p^{m-1}, p^n}, \tag{4.1}$$

provided $p > 2$ and $(p-1) \nmid m$. Additional results of this kind can be stated without any difficulty.

It is perhaps of greater interest to point out one or two sequences of a more recondite sort that also satisfy the hypotheses of our theorems. Let $p > 2$ and let l be a rational number that is integral (mod p). Let $sn\, x = sn(x, l)$ denote the Jacobi elliptic function with modulus $k^2 = l$. Put

$$sn\, x = \sum_0^\infty A_{2m+1} x^{2m+1}/(2m+1)! \qquad (c_1 = 1) \tag{4.2}$$

and

$$x/sn\, x = \sum_0^\infty \beta_{2m} x^{2m}/(2m)! \qquad (\beta_0 = 1). \tag{4.3}$$

Then [1] the sequence $\{A_m\}$ satisfies $A^m (A^{p-1} - A_p)^r \equiv 0 \pmod{p^r, p^m}$; also the sequence $\{\eta_m\} = \{\beta_m/m\}$ satisfies $\eta^m (\eta^{p-1} - A_p)^r \equiv 0 \pmod{p^r, p^{m-1}}$ provided $(p-1) \nmid m$. Thus our results hold for the sequences $\{A_m\}$ and $\{\eta_m\}$. In particular we can state such congruences as

$$A_1^{m_1} \cdots A_k^{m_k} (\lambda_1 A_1 + \cdots + \lambda_k A_k)^r \equiv 0 \pmod{p^r, p^{m_1}, \cdots, p^{m_k}} \tag{4.4}$$

and

$$\eta_1^{m_1} \cdots \eta_k^{m_k} (\lambda_1 \eta_1 + \cdots + \lambda_k \eta_k)^r \equiv 0 \pmod{p^r, p^{m_1-1}, \cdots, p^{m_k-1}}, \tag{4.5}$$

where in (4.5) it is assumed that $(p-1) \nmid m_i$. Also corresponding to (4.1) we can state congruences of the type

$$A^m \eta^n (A^{p-1} - \eta^{p-1})^r \equiv 0 \pmod{p^r, p^m, p^{n-1}}, \tag{4.6}$$

where $(p-1) \nmid n$. In the second place we may vary the modulus l. We accordingly consider k sequences $\{A_m(l_i)\}$ defined by means of

$$sn(x, l_i) = \sum_0^\infty A_{2m+1}(l_i) x^{2m+1}/(2m+1)! \qquad (i = 1, \cdots, k);$$

we define $\beta_m(l_i)$ in the obvious way. Then Theorem 1 yields

$$A^{m_1}(l_i)\cdots A^{m_k}(l_k)\,(\lambda_1 A^{u_1}(l_i)C(l_1)+\cdots+\lambda_k A^{u_k}(l_k)C(l_k))^r\equiv 0$$

(mod $p^{re}, p^{m_1}, \cdots, p^{m_k}$), where the u_i satisfy (3.3) and as in (3.4) $C(l_i)=\prod_{j=1}^{k}{}' A^{u_i/(p-1)}(l_i)$, where $j\neq i$. Similarly

$$\eta_1^{m_1}(l_1)\cdots\eta^{m_k}(l_k)\,(\lambda_1\eta^{u_1}(l_1)C(l_1)+\cdots+\lambda_k\eta^{u_k}(l_k)C(l_k))^r\equiv 0$$

(mod $p^{re}, p^{m_1-1}, \cdots, p^{m_k-1}$), where now $(p-1)\nmid m_i$.

As a special case of these results concerning the coefficients of the Jacobi functions we may mention the coefficients of the lemniscate function studied by Hurwitz [4].

DUKE UNIVERSITY.

REFERENCES.

[1] L. Carlitz, "Congruences for the coefficients of the Jacobi elliptic functions," *Duke Mathematical Journal*, vol. 16 (1949), pp. 297-302.

[2] ———, "Kummer congruences and the Schur derivative," *American Journal of Mathematics*, vol. 75 (1953), pp. 699-706.

[3] ———, "Some theorems on Kummer's congruences," *Duke Mathematical Journal*, vol. 20 (1953), pp. 423-431.

[4] A. Hurwitz, "Über die Entwickelungs coeffizienten der lemniscatischen Functionen," *Mathematische Annalen*, vol. 51 (1899), pp. 196-226 (= *Mathematische Werke*, Basel, 1933, vol. 2, pp. 342-373).

[5] H. S. Vandiver, "Note on a certain ring congruence," *Bulletin of the American Mathematical Society*, vol. 43 (1937), pp. 418-423.

ON THE FUNCTIONAL EQUATION $dy/dx = f(x, y(x), y(x+h)), h > 0$.*

By Shafik Doss and Saad K. Nasr.

1. The object of the present note is to show that if certain conditions bearing on the function $f(x, y, z)$ of three variables are satisfied, then there is just one bounded solution of the given functional equation, defined for $x \geqq x_0$ and satisfying the initial condition $y(x_0) = y_0$.

To prove the statement we shall apply Picard's method of successive approximations, making the following assumptions which will insure the convergence of the process:

(i) $|f(x, Y, Z) - f(x, y, z)| \leqq K_1(x)|Y - y| + K_2(x)|Z - z|$,

(ii) $\int_{x_0}^{\infty} (K_1(x) + K_2(x))\,dx = B < 1$,

(iii) there exist a y and a z for which $\int_{x_0}^{\infty} |f(t, y, z)|\,dt < \infty$.

Condition (i) is a Lipschitz hypothesis of a very strong type. If the integral in (ii) is convergent, then condition (ii) can always be satisfied by choosing x_0 large enough. The solution will be defined for $x \geqq x_0$.

With reference to a constant y_0, consider the sequence of functions defined by

$$y_1(x) = y_0 + \int_{x_0}^{x} f(t, y_0, y_0)\,dt,$$

$$y_n(x) = y_0 + \int_{x_0}^{x} f(t, y_{n-1}(t), y_{n-1}(t+h))\,dt \qquad n \geqq 2.$$

We have $|y_1(x) - y_0| \leqq \int_{x_0}^{x} |f(t, y_0, y_0)|\,dt \leqq A$, where A is finite by (i), (ii) and (iii). If we assume

$$|y_{p-1}(x) - y_{p-2}(x)| \leqq AB^{p-2}, \qquad 2 \leqq p \leqq n, \qquad y_0(x) = y_0,$$

* Received December 15, 1952; revised April 30, 1953.

then

$$|y_n(x) - y_{n-1}(x)| \leqq \int_{x_0}^{x} \{K_1(t)|y_{n-1}(t) - y_{n-2}(t)| + K_2(t)|y_{n-1}(t+h) - y_{n-2}(t+h)|\}dt \leqq AB^{n-1}.$$

The series $\Sigma(y_n(x) - y_{n-1}(x))$ is therefore absolutely and uniformly convergent and the function $y(x) = \lim_{n\to\infty} y_n(x)$ exists and satisfies

$$y(x) = y_0 + \int_{x_0}^{x} f(t, y(t), y(t+h))dt, \tag{1}$$

which is equivalent to the given functional equation with the initial condition $y(x_0) = y_0$. We observe that $|y(x)| \leqq |y_0| + A/(1-B)$.

It is now possible to prove that $y(x)$ is the only *bounded* solution $Y(x)$ which satisfies (1). In fact, if $Y(x)$ is such a solution of (1), then

$$Y(x) = y_0 + \int_{x_0}^{x} f(t, Y(t), Y(t+h))dt,$$

and we can easily deduce, with the previous notations and in virtue of conditions (i) and (ii), that

$$|Y(x) - y_n(x)| \leqq MB^n, n \geqq 1, \text{ where } M = \max_t \{|Y(t) - y_0|\}.$$

We therefore get $Y(x) = \lim_{n\to\infty} y_n(x) = y(x)$ and our statement is proved.

If we now assume that condition (i) is satisfied for Y, y, Z, z belonging to the interval $(y_0 - b, y_0 + b)$, then, provided $A/(1-B) < b$ holds (and this is always possible if x_0 is taken large enough), we get by induction

$$|y_n(x) - y_{n-1}(x)| \leqq AB^n \text{ and } |y_n(x) - y_0| < A/(1-B) < b.$$

Therefore, the solution $y(x) = \lim_{n\to\infty} y_n(x)$ exists and satisfies

$$|y(x) - y_0| < b, \tag{2}$$

and just as before we may show that this is the only solution satisfying (2).

Particular case. For the equation

$$dy/dx = A(x)y(x) + B(x)y(x+h), h > 0, \tag{3}$$

the previous considerations will apply if we assume that

$$\int^{\infty} |A(x)|\,dx \text{ and } \int^{\infty} |B(x)|\,dx \text{ are convergent.} \tag{4}$$

Hence (3) has, under the assumptions (4), just one bounded solution defined for $x \geqq x_0$ and satisfying the initial condition $y(x_0) = y_0$.

2. We owe to a referee the following corollaries and comments.

(a) If $y(x)$ is a bounded solution, then $\lim y(x) = \lambda$ exists as $x \to \infty$.

(b) The limit λ can have any real value, and the correspondence $y \leftrightarrow \lambda$ is one-to-one.

(c) If $x_0 = -\infty$, there still is a one-parameter family of bounded solutions; for every bounded solution both $y(-\infty)$ and $y(+\infty)$ exist; each one of these can be chosen arbitrarily and determines the solution uniquely; the correspondence $y(-\infty) \leftrightarrow y(+\infty)$ is one-to-one again.

(d) If the integral in (ii) is not convergent, the theorem is no longer true.

(e) The case of the linear homogeneous equation (3) is a consequence of the following stronger theorem (which seems to be classical):

If $\int_0^\infty |K(x,t)|\,dt \leqq C < 1$, then $y(x) = y_0 + \int_0^\infty K(x,t)y(t)\,dt$ has just one bounded solution.

To show that, for a bounded solution $y(x)$, the limit $y(\infty)$, exists, consider $y(x)$ as given by (1) and let $|y(x)| < M$. We have, in virtue of (i), (ii) and (iii),

$$\int_{x_0}^{x} |f(t, y(t), y(t+h))|\,dt \leqq \int_{x_0}^{x} |f(t,y,z)|\,dt + |M-y| \int_{x_0}^{x} K_1(t)\,dt + |M-z| \int_{x_0}^{x} K_2(t)\,dt < A,$$

where A is finite and independent of x. This shows that

$$\int_{x_0}^{\infty} f(t, y(t), y(t+h))\,dt$$

is convergent and $\lambda = \lim_{x\to\infty} y(x)$ exists and is finite.

To show (b), consider a bounded solution $Y(x)$ of

$$Y(x) = Y_0 + \int_{x_0}^{x} f(t, Y(t), Y(t+h))\,dt.$$

It is easy to see that

$$(5) \qquad |Y_0 - y_0|/(1+B) \leqq |Y(x) - y(x)| \leqq |Y_0 - y_0|/(1-B).$$

If we observe that, in virtue of our uniqueness theorem, λ is a non-decreasing function of y_0, we see that (5) implies property (b).

The previous considerations apply if $x_0 = -\infty$ and property (c) is therefore true.

To prove property (d), consider

$$y'(x) = (y(x) - y(x+h))/(e^{-h} - 1). \tag{6}$$

A solution of (6) is $y(x) = e^{-x}$. For $x \geqq x_0$, the functions $y(x) = ae^{-x} + b$ form a family of bounded solutions depending on two arbitrary constants. There is thus an infinity of bounded solutions satisfying the initial condition $y(x_0) = y_0$.

To show (e), we only have to define $K(x, t)$, for $x \geqq 0$, $t \geqq 0$, as follows:

$$K(x,t) = A(t) \text{ for } x \geqq t,\ t < h;\ K(x,t) = A(t) + B(t-h) \text{ for } x \geqq t \geqq h;$$

$$K(x,t) = B(t-h) \text{ for } 0 \leqq t-h < x < t;\ K(x,t) = 0 \text{ for } x < t < h;$$

$$K(x,t) = 0 \text{ for } x \leqq t-h.$$

Then any bounded solution of (3) satisfying the initial condition $y(0) = y_0$ will be a bounded solution of

$$y(x) = y_0 + \int_0^\infty K(x,t)y(t)\,dt. \tag{7}$$

That (7) has, under the assumption $\int_0^\infty |K(x,t)|\,dt \leqq C < 1$, only one bounded solution, can be seen, just as in the proof of our theorem, by the consideration of the sequence of functions

$$y_1(x) = y_0 + \int_0^\infty K(x,t)y_0\,dt, \quad y_n(x) = y_0 + \int_0^\infty K(x,t)y_{n-1}(t)\,dt, \quad n \geqq 2.$$

Property (e) is therefore proved.

FACULTY OF SCIENCE,
ALEXANDRIA UNIVERSITY, EGYPT.

ON NON-OSCILLATORY LINEAR DIFFERENTIAL EQUATIONS.*

By PHILIP HARTMAN and AUREL WINTNER.

There are various senses in which the linear differential equation

$$x'' + f(t)x = 0, \tag{1}$$

where $f(t)$ is a given continuous function defined for large positive t, can be required to have the same asymptotic behavior $(t \to \infty)$ as the case $f(t) \equiv 0$ of (1). First, since $x'' + 0x = 0$ has the two solutions $x = x(t) \equiv 1$, $x = y(t) \equiv t$, the simplest condition under which $f(t)$ may be considered to be *small* $(t \to \infty)$ in (1) appears to be the following requirement: (1) has some pair of solutions $x = x(t)$, $y = y(t)$ satisfying

$$(2_1) \qquad x(t) \to 1; \qquad\qquad (2_2) \qquad y(t) \sim t,$$

as $t \to \infty$. A stronger requirement is the existence of a pair of solutions for which these asymptotic relations hold and for which the derivatives satisfy

$$(3_1) \qquad x'(t) \to 0; \qquad\qquad (3_2) \qquad y'(t) = o(t).$$

A still stronger requirement is the pair of relations

$$(4_1) \qquad x'(t) = o(t^{-1}); \qquad\qquad (4_2) \qquad y'(t) \to 1,$$

the second of which, being the result of a formal differentiation of (2_2), is a natural desideratum, whereas (4_1), being a substantial refinement of the formal derivative (3_1) of (2_1), appears to be quite artificial. But this appearance is misleading, since it turns out that

(5) the existence of a solution $y = y(t)$ satisfying (4_2) is equivalent to the existence of a solution $x = x(t)$ satisfying (4_1). In addition,

(6) the existence of a solution $y = y(t)$ satisfying both (3_2) and (2_2) is equivalent to the existence of a solution satisfying both (3_1) and (2_1). Finally,

(7) the existence of a solution $x = y(t)$ satisfying (2_2) is equivalent to the existence of a solution $x = x(t)$ satisfying (2_1). Since it is clear (by integration) that

(8) condition (4_2) [but not (3_2)] implies (2_2),

---

* Received March 5, 1953.

it follows that *the assumptions* (3_2), (4_1) *are equivalent to the formal differentiability of the respective asymptotic relations* (2_1), (2_2). Note that, in view of (7) and of the superposition principle, (1) can never have two solutions satisfying (2_1), simply because exactly one of the two solutions (2_1), (2_2) is bounded as $t \to \infty$ (so that any two such solutions are linearly independent).

In order to prove (5) and (6), grant first (7) and note that, since (2_1) and (2_2) are two linearly independent solutions of (1), their Wronskian is a non-vanishing constant,

$$x(t)y'(t) - y(t)x'(t) = \text{const.} \neq 0. \tag{9}$$

Both assertions of (5) and both assertions of (6) follow immediately if each of the four pairs of asymptotic relations assumed in (5) and (6) is substituted into (9).

In order to prove (7), note that, if $x = u(t)$ is any non-vanishing solution of (1), then direct differentiations show that $x = v(t)$ and $x = w(t)$, where

$$(10_1) \quad v(t) = u(t) \int^t (u(s))^{-2} ds; \qquad (10_2) \quad w(t) = u(t) \int_t^\infty (u(s))^{-2} ds,$$

are also solutions of (1), with the understanding that $w(t)$ exists only if the integral occurring in (10_2) is convergent. The assertion of (7) follows if (10_1) is applied to $u(t) = x(t)$ or (10_2) to $u(t) = y(t)$ according as (2_1) or (2_2) is assumed.

This proves that the case $k = 1$ of any of the three properties (2_k), $(2_k) + (3_k)$, $(2_k) + (4_k)$ of (1) or $f(t)$ implies and is implied by the corresponding case $k = 2$. Hence it is sufficient to consider each of these three properties for the case $k = 1$ alone. The first of these three properties consists of the existence of a solution curve $x = x(t)$ which, as $t \to \infty$, tends to a line λ parallel to, but distinct from, the t-axis of the (t, x)-plane; cf. (2_1). The second and third properties require, besides the first property, that, as $t \to \infty$, the line λ should become a limit tangent and an asymptote, respectively; cf. (3_1) and (4_1). In fact, the definition of a curve $x = x(t)$ having an asymptote is the existence of a finite limit

$$\lim_{t \to \infty} (x(t) - tx'(t)), \tag{11}$$

which, if (2_1) or the existence of a finite $x(\infty) = \lim x(t)$ is assumed, is equivalent to (4_1).

It will be convenient to introduce the following terminology: (1) or $f(t)$ will be said to have the property (2), (3) or (4) if (1) has solutions $x = x(t)$, $x = y(t)$ satisfying (2), (2) and (3), or (2) and (4), respectively.

In what follows, sufficient conditions, and also necessary conditions, will be proved for an f having property (4), and also for an f having the weaker property (3). On the other hand, there will result no conditions which are sufficient for property (2) without being sufficient for (3) too.

(i) *In order that* (1) *has property* (3), *it is sufficient that the improper integral*

$$\int^{\infty} f(t)\,dt \text{ converges} \qquad \left(\int^{\infty} = \lim_{T\to\infty} \int^{T}\right) \tag{12}$$

(*possibly just conditionally*) *and that the corresponding indefinite integral*

$$F(t) = \int_{t}^{\infty} f(s)\,ds \qquad \left(\int_{t}^{\infty} = \lim_{T\to\infty} \int_{t}^{T}\right) \tag{13}$$

satisfies the following conditions:

$$\int^{\infty} F(t)\,dt \text{ converges} \qquad \left(\int^{\infty} = \lim_{T\to\infty} \int^{T}\right) \tag{14}$$

(*possibly just conditionally*) *and*

$$\int^{\infty} tF^2(t)\,dt < \infty. \tag{15}$$

The following partial converse of (i) will be clear from the proof of (i):

(i bis) *If conditions* (12)-(14) *are satisfied, then* (15) *is not only sufficient but necessary as well in order that* (1) *has property* (3).

REMARK. It follows from (i bis) that (14) and (15) in (i) cannot be replaced by

$$\int^{\infty} |F(t)|\,dt < \infty. \tag{16}$$

The following counterpart of (i) will also be proved:

(ii) *In order that* (1) *has property* (4), *it is sufficient that the conditions* (12)-(15) *of* (i) *hold and that*

$$F(t) = o(t^{-1}) \qquad (t \to \infty). \tag{17}$$

What now corresponds to (i bis) is the following fact:

(iii) *In order that* (1) *possesses property* (4), *it is necessary that* (12) *holds and that the function* (13) *satisfies* (17).

REMARK. If (12) is satisfied by (1), that is, if the function $F(t)$ exists (which, according to (iii), is necessary for property (4)), and *if* $F(t)$ *does not change sign from a certain* $t = t_0$ *onward* (which is the case if, but not only if, $f(t)$ does not change sign from a certain $t = t^0$ onward), *then not only* (17) *but also* (16) *is necessary in order that* (1) *has property* (4). This can be seen as follows: According to (1) and (13),

$$(x - tx')' \equiv -tx'' = tfx = -txF' \equiv (-txF)' + (x + tx')F.$$

If a quadrature is applied to this identity, then, since (4) and (17) imply that $-txF = o(1)$, it follows that the limit (11) exists if and only if the integral

$$\int^{\infty} (x + tx')F dt \text{ converges} \qquad \left(\int^{\infty} = \lim_{T\to\infty} \int^{T}\right).$$

In view of (4), this proves the last italicized assertion.

A set of sufficient conditions, different from those supplied by (i)-(ii), for (1) to possess the property (4) is contained in the following theorem:

(iv) *In order that* (1) *has property* (4), *it is sufficient that*

$$\int^{\infty} tf(t)dt \text{ converges} \qquad \left(\int^{\infty} = \lim_{T\to\infty} \int^{T}\right) \tag{18}$$

and

$$\int^{\infty} t^{2p-1} |f(t)|^p dt < \infty, \tag{19}$$

where p *is some index on the range*

$$1 \leqq p \leqq 2. \tag{20}$$

The assumptions (12)-(15) and (17) of (i) and (ii) do not require that

$$\int^{\infty} |f(t)|\, dt < \infty, \tag{21}$$

nor that (16) holds, and still less that

$$\int^{\infty} \max_{t\leqq s<\infty} |F(s)|\, dt < \infty \tag{22}$$

or (what is still more stringent) that

$$\int^{\infty} t\,|f(t)|\,dt < \infty. \tag{23}$$

On the other hand, (23) or (12) and (22) imply the conditions (12)-(15) and (17) of (i) and (ii). In fact, both (22) and (23) imply (22) (hence (16)) and (17), while (16) and (17) imply (15). Thus (i) improves a classical result (cf. [5], p. 486, footnotes 57 and 58, and [8], p. 854, footnote), in which (23) is assumed, and a refinement of this result ([7], p. 595), in which (12) and (22) are assumed, and in both of which results it is asserted that (1) has the property (3). Actually, the literature consulted does not seem to point out, even under the assumption (23), that the conclusion of (i) can be strengthened to that of (ii). It is, of course, easy to see that (1), (2_1) and (23) imply (4_1).

In the course of the proof of (i), where it will be assumed that (12) holds, it will be convenient to introduce the functions

$$G(t) = H^2(t)\int^{t} ds/H^2(s), \text{ where } H(t) = \exp\int^{t} F(s)\,ds. \tag{24}$$

Since (14) implies that

$$H(t) \to \text{const.}\ (>0) \text{ and } G(t) \sim t \qquad (t \to \infty), \tag{25}$$

it is clear that (i), (i bis) and (ii) are contained in the following theorem:

(v) *If* (12) *is assumed, and if* $G(t)$ *and* $H(t)$ *are defined by* (24), *then* (1) *possesses a solution* $x = x(t)$ *and/or a solution* $x = y(t)$ *satisfying*

$$(26_1) \quad x \sim H(t); \qquad (26_2) \quad y \sim G(t)/H(t)$$

$(t \to \infty)$ *if and only if*

$$\int^{\infty} dt/H^2(t) = \infty \tag{27}$$

and

$$\int^{\infty} G(t)F^2(t)\,dt < \infty; \tag{28}$$

in which case the relations

$$(29_1) \quad x'/x = F + o(G^{-1}); \qquad (29_2) \quad y'/y = F + (1 + o(1))G^{-1}$$

also hold.

REMARK. It follows from a criterion of [9], pp. 376-377, that *assumption* (27) *of* (v) *is satisfied whenever* (1) *is non-oscillatory.*

(iv) is contained in Theorems (vi) and (vii) below.

(vi) *In order that* (1) *posssesses a solution* $x = x(t)$ *and/or a solution* $x = y(t)$ *satisfying*

$$(30_1) \qquad x'/x = o(t^{-1}); \qquad (30_2) \qquad y'/y \sim t^{-1}$$

$(t \to \infty)$, *it is necessary and sufficient that*

$$(31) \qquad \text{l. u. b.}_{0<s<\infty} \left| \int_t^{t+s} rf(r)\,dr \right| / (1 + \log(1 + s/t)) \to 0 \quad \text{as} \quad t \to \infty.$$

It is easily verified that (31) is satisfied if either (18) holds or (19) holds for some $p \geqq 1$. It is understood that (vi) is meant to imply that (31) is sufficient for (1) to be non-oscillatory (that is, in order that no solution $x = x(t) \not\equiv 0$ of (1) can vanish for large t-values).

(vii) *If* (19) *and* (20) *hold, then* (1) *possesses a pair of solutions* $x = x(t)$, $x = y(t)$ *satisfying*

$$(32_1) \qquad x \sim \exp \int^t sf(s)\,ds; \qquad (32_2) \qquad y \sim t \exp - \int^t sf(s)\,ds$$

and (30_1), (30_2), *respectively.*

This furnishes a criterion for (1) to have the property (2), as well as for the following situation: (1) has solutions $x = x(t)$, $y = y(t)$ satisfying

$$(33_1) \qquad 0 \not\equiv x = o(1); \qquad (33_2) \qquad yt^{-1} \to \infty.$$

COROLLARY. *If* (19) *and* (20) *hold and if*

$$(34) \qquad \int^t sf(s)\,ds \to -\infty, \qquad (t \to \infty),$$

then (1) *has solutions* $x = x(t) \not\equiv 0$ *and* $x = y(t)$ *which satisfy* (33_1), (30_1) *and* (33_2), (30_2), *respectively.*

The above theorems will be proved in the following order: (v), (vi), (vii), (iii).

Proof of (v). With a fixed choice of the lower limit of integration, put

$$(35) \qquad z = z(t) = \exp \int^t F(s)\,ds.$$

Then $z(t)$ satisfies the differential equation

$$z'' + (f - F^2)z = 0. \tag{36}$$

Thus, if $x = x(t)$ is a solution of (1) and if $u = u(t)$ is defined by

$$x = zu, \tag{37}$$

then u satisfies the differential equation

$$(z^2u')' + F^2z^2u = 0; \tag{38}$$

conversely, if $u = u(t)$ is a solution of (38), then (37) is a solution of (1). Thus, (1) has a solution $x = x(t)$ satisfying (26_1) if and only if (38) has a solution satisfying

$$u \to 1 \qquad (t \to \infty). \tag{39}$$

Introduce the new independent variable

$$\tau = \int^t z^{-2}(s)\,ds = \int^t (\exp -2\int^s F(r)\,dr)\,ds \tag{40}$$

and let $\dot{u}$ denote $du/d\tau$. Then (27) means that $\tau \to \infty$ as $t \to \infty$. The differential equation (38) becomes

$$\ddot{u} + F^2z^4u = 0 \qquad (t = t(\tau)). \tag{41}$$

Since $F^2z^4 \geqq 0$, a necessary and sufficient condition in order that (41) has a solution $u = u(\tau)$ satisfying (39) as $\tau \to \infty$ (or $t \to \infty$) is that

$$\int^\infty \tau F^2z^4\,d\tau < \infty; \tag{42}$$

cf. [7], Appendix, pp. 601-602. (The condition (42) on (41) is the analogue of condition (23) on (1).) In view of (24), (35) and (40), the inequality (42) is precisely (28).

When (28), that is, (42), holds, then (41) has a pair of solutions $u = u(\tau)$, $u = v(\tau)$ satisfying

$$(43_1) \quad u \to 1, \quad \dot{u} = o(\tau^{-1}); \qquad (43_2) \quad v \sim \tau, \quad \dot{v} \to 1,$$

as $\tau \to \infty$. Furthermore, the existence of one of these solutions is equivalent to the existence of the other; cf. (5).

Consider the solutions $x = zu$ and $y = zv$ of (1). Clearly, (26_1) and (26_2) hold. Since $x'/x = z'/z + \tau'\dot{u}/u$ and $y'/y = z'/z + \tau'\dot{v}/v$, it follows from (40) and (43), that (29_1) and (29_2) hold. Thus, if (12) is assumed,

(27) and (28) are sufficient for the existence of a solution $x = x(t)$ of (1) satisfying (26_1), (29_1) and/or for the existence of a solution $x = y(t)$ satisfying (26_2), (29_2).

There remains to be proved the necessity of (27) and (28). If (1) has a solution satisfying (26_1) or (26_2), then (1) is non-oscillatory. This implies (27) (cf. [9], pp. 376-377; for a refinement of this assertion, cf. [3]). Finally, in the above proof of the sufficiency of (28), it was pointed out that (28) is necessary and sufficient when (12) and (27) hold. This completes the proof of (v).

Proof of (vi) *and* (vii). Consider the variation of constants and the change of independent variables

$$x = e^{\tau} u \tag{44}$$

and

$$\tau = \tfrac{1}{2} \log t \qquad (t = e^{2\tau}), \tag{45}$$

respectively; so that

$$\tau' = \tfrac{1}{2} t^{-1} = \tfrac{1}{2} e^{-2\tau}; \qquad \dot{t} = 2t = 2e^{2\tau}, \tag{46}$$

if $\dot{u} = du/d\tau$. By virtue of (44) and (45), the differential equation (1) is equivalent to

$$\ddot{u} - (1 + \phi(\tau)) u = 0, \tag{47}$$

where

$$\phi(\tau) = -4t^2 f(t), \qquad t = e^{2\tau}. \tag{48}$$

A necessary and sufficient condition that (1) be non-oscillatory and possess a solution $x = x(t)$ $[x = y(t)]$ satisfying (30_1) $[(30_2)]$ is that (47) be non-oscillatory and have a solution $u = u(\tau)$ $[u = v(\tau)]$ satisfying

$$\dot{u} \sim -u; \qquad (49_1) \qquad\qquad \dot{v} \sim v. \qquad (49_2)$$

This is clear if $x = e^{\tau} u$ and $y = e^{\tau} v$, since $x'/x = \tau'(1 + \dot{u}/u)$ and $y'/y = \tau'(1 + \dot{v}/v)$, where $\tau' = \frac{1}{2} t^{-1}$. On the other hand, a necessary and sufficient condition for (47) to be non-oscillatory and to have a solution $u = u(\tau)$ and/or a solution $u = v(\tau)$ satisfying (49_1) or (49_2), respectively, is that

$$\text{l. u. b.}_{0<\sigma<\infty} \Big| \int_{\tau}^{\tau+\sigma} \phi(\rho)\, d\rho \Big| / (1 + \int_{\tau}^{\tau+\sigma} d\rho) \to 0, \tag{50}$$

as $\tau \to \infty$; cf. [2], Theorem (IV), p. 570. Since (50) is equivalent to (31) by virtue of (48), assertion (vi) follows.

In order to prove (vii), note that (19) is equivalent to

$$\int^{\infty} |\phi(\tau)|^p \, d\tau < \infty, \tag{51}$$

by (48). By [2], Theorem (VII), p. 575, condition (51) assures the existence of solutions $u = u(\tau)$, $v = v(\tau)$ of (47) satisfying, as $\tau \to \infty$,

$$u(\tau) \sim \exp(-\tau - \tfrac{1}{2}\int^{\tau} \phi(\sigma)\, d\sigma)\,; \tag{52$_1$}$$

$$v(\tau) \sim \exp(\tau + \tfrac{1}{2}\int^{\tau} \phi(\sigma)\, d\sigma) \tag{52$_2$}$$

and (49$_1$), (49$_2$). The corresponding solutions $x = e^{\tau}u$ and $y = e^{\tau}v$ of (1) satisfy (30) and (32). This proves (vii).

Reversing the procedures of the proofs of (vi) and (vii), it is possible to obtain conclusions concerning the asymptotic behavior of solutions of (47) from those concerning solutions of (1). For example, (i) and (v) supply the following criterion:

Let $\phi(\tau)$ be continuous for large τ and have the properties that

$$\int^{\infty} \phi(\tau) e^{-2\tau} d\tau \text{ converges} \qquad \left(\int^{\infty} = \lim_{T\to\infty} \int^{T}\right) \tag{53}$$

(possibly just conditionally) and that the corresponding indefinite integral

$$\Phi(\tau) = \int_{\tau}^{\infty} \phi(\sigma) e^{-2\sigma} d\sigma \qquad \left(\int^{\infty} = \lim_{T\to\infty} \int^{T}\right) \tag{54}$$

satisfies the following conditions:

$$\int^{\infty} \Phi(\tau)\, d\tau \text{ converges} \qquad \left(\int^{\infty} = \lim_{T\to\infty} \int^{T}\right) \tag{55}$$

(possibly just conditionally) and

$$\int^{\infty} e^{4\tau}\Phi^2(\tau)\, d\tau < \infty. \tag{56}$$

Then (47) possesses a pair of solutions $u = u(\tau)$, $v = v(\tau)$ satisfying

$$u \sim e^{-\tau}; \tag{57$_1$}$$

$$v \sim e^{\tau} \tag{57$_2$}$$

and

$$\dot{u}/u = -1 + O(|\Phi(\tau)|\, e^{2\tau}) + o(1); \tag{58_1}$$

$$\dot{v}/v = 1 + O(|\Phi(\tau) e^{2\tau}) + o(1). \tag{58_2}$$

Proof. By (45) and (48), the relations (12), (14), (15) are respectively identical with (53), (55), (56); while, by (44), the existence of solutions $x = x(t)$, $x = y(t)$ of (1) satisfying (3) is equivalent to the existence of solutions of (47) satisfying (57). Finally, (58) follows from (29) by virtue of (44), (45).

Similarly, (v) itself leads to theorems concerning asymptotic integrations of (47) in which (57), the analogue of (2), is replaced by

$$u \sim \exp(-\tau + \int^{\tau} \Phi(\sigma)\, d\sigma); \tag{59_1}$$

$$v \sim \exp(-\tau + \int^{\tau} \Phi(\sigma)\, d\sigma) \cdot \int^{\tau} (\exp -2 \int^{\sigma} \Phi(\rho)\, d\rho)\, d\sigma. \tag{59_2}$$

Proof of (iii). If a solution $x = x(t)$ of (i) does not vanish (on a t-interval), then $l = x'/x$ exists and satisfies the Riccati differential equation $l' + l^2 + f(t) = 0$ (on that t-interval). If $x = x(t)$ satisfies (2_1), then

$$l(t) + \int^{t} l^2(s)\, ds + \int^{t} f(s)\, ds = l(t_0), \tag{60}$$

if the lower limit of integration $t = t_0$ is sufficiently large and fixed. If, in addition, $x = x(t)$ satisfies (4_1), then $l(t) \to 0$, and the first integral in (60) tends to a finite limit, as $t \to \infty$. Hence (12) holds. Consequently, (60) can be written as

$$l(t) = \int_{t}^{\infty} l^2(s)\, ds + F(t). \tag{61}$$

In view of (4_1), this implies that $o(t^{-1}) = o(t^{-1}) + F(t)$, which proves (13).

This completes the proofs of all assertions made above.

THE PROPERTY (4*). In what follows, differential equations (1) will be considered for which the existence of a solution $x = y(t)$ satisfying (4_2) can be improved to

$$y(t) = t + o(1), \qquad y'(t) = 1 + o(t^{-1}). \tag{62}$$

It turns out that this is the case if and only if *every solution of* (1) *has an asymptote*, that is, if and only if (1) has two linearly independent solutions $x = x_1(t)$, $x = x_2(t)$ each of which leads to a finite limit (11). Then (1) will be said to have *property* (4*). It will be seen below that property (4) is necessary, though not sufficient, for property (4*).

The following remarks (α), (β) will be needed:

(α) If $x = x(t)$ is any fixed solution of (1), it has an asymptote if and only if the improper integral

$$\int^{\infty} tf(t)x(t)\,dt \text{ converges} \qquad \left(\int^{\infty} = \lim_{T\to\infty} \int^{T}\right). \tag{63}$$

(β) A function $x = x(t)$, possessing a continuous first derivative $x'(t)$ for large positive t, has an asymptote if and only if there exists a pair of constants a, b satisfying both conditions

$$(64_{\mathrm{I}}) \quad x(t) = at + b + o(1); \qquad (64_{\mathrm{II}}) \quad x'(t) = a + o(t^{-1}) \qquad (t \to \infty).$$

Proof of (α). Since (1) means that $(x - tx')' = -tx''$ is identical with tfx, the existence of the limit (11) is equivalent to condition (63).

Proof of (β) (cf. [1], pp. 144-145). The existence of the limit (11) means that $x = x(t)$ satisfies a linear differential equation of the form $x = tx' + \text{const.} + h(t)$, where $h(t)$ is a (continuous) function satisfying $h(t) \to 0$ as $t \to \infty$. If this linear differential equation of first order is integrated by a quadrature, (64_{I}) follows. But direct substitutions show that if $x(t)$ is of the form (64_{I}), then (64_{II}) is equivalent to the existence of the limit (11).

It is easy to conclude from (β) that property (4*) implies property (4). In fact, if there belongs to every solution $x(t)$ of (1) a pair of constants a, b satisfying (64_{I}) and (64_{II}), then $a \not\equiv 0$ for some solution. For, if it is assumed that (64_{I}) and (64_{II}) hold for two linearly independent solutions $x = x_1(t)$, $x = y_2(t)$ with $(a, b) = (0, b_1)$, $(a, b) = (0, b_2)$, respectively, then the Wronskian $x_1x_2' - x_2x_1'$ is $O(1)o(1) + O(1)o(1) = o(1)$, which is impossible, since the Wronskian of two linearly independent solutions of (1) is a non-vanishing constant. Accordingly, property (4*) implies the existence of a solution $x = y(t)$ for which (64_{I}) and (64_{II}) hold with $a = 1$. But this implies (4_2), and therefore the existence of a solution $x = x(t)$ satisfying (2_1) and (4_1).

The equivalence of property (4*) and of the existence of a solution

$y = y(t)$ satisfying (62) now follows by taking a linear combination of solutions satisfying (2_1)-(4_1) and the case $a = 1$ of (64_{I})-(64_{II}).

In particular, (1) must possess a solution $x(t) \sim t$. It follows that *if* $f(t)$ *does not change sign from a certain* $t = t^0$ *onward* (that is, if

$$f(t) = \pm\, |\, f(t)\,|, \tag{65}$$

where the choice of the sign is independent of $t \geqq t^0$), *then the condition*

$$\int^{\infty} t^2\, |\, f(t)\,|\, dt < \infty \tag{66}$$

is necessary and sufficient in order that (1) *has property* (4*). In fact, if (1) has property (4*), then (63) must be satisfied by every solution $x(t)$, and therefore by some solution $x(t) \sim t$, and so the necessity of (66) follows from the assumption (65). On the other hand, the sufficiency of (66) for the (4*)-character of (1) is true without the assumption (65) also. This follows as a particular case ($g \equiv 0$) of the following criterion (γ):

(γ) If $f(t)$ and $g(t)$ are continuous functions satisfying (66) and

$$\int^{\infty} t\, |\, g(t)\,|\, dt < \infty, \tag{67}$$

then every solution $x = x(t)$ of

$$x'' + g(t)x' + f(t)x = 0 \tag{68}$$

has an asymptote.

This criterion (γ) is known; cf. [4], where, however, a lengthy proof is given, since it is not observed that (γ) is an easy consequence of the following lemma (γ bis) which, in another form, is of a much older date (cf. [5], p. 486, footnotes 57 and 58, and [8], p. 854, footnote; for a simple proof, cf. [6], pp. 262-264).

(γ bis) Let the coefficients $a_{ik}(t)$ of a linear differential system

$$u' = a_{11}(t)u + a_{12}(t)v, \qquad v' = a_{21}(t)u + a_{22}(t)v \tag{69}$$

be continuous functions satisfying

$$\int^{\infty} |\, a_{ik}(t)\,|\, dt < \infty \qquad (i, k = 1, 2). \tag{70}$$

Then there belongs to every pair of constants α, β a unique solution of (69)

satisfying $(u(t), v(t)) \to (\alpha, \beta)$ as $t \to \infty$ (which implies that the limits $\alpha = u(\infty)$, $\beta = v(\infty)$ exist for every solution).

In order to deduce (γ) from (γ bis), put, with reference to any solution $x = x(t)$ of (68),

$$u = x - tx', \qquad v = x'. \tag{71}$$

Then $u' = -tx'' = t(gx' + fx) = tgv + tf(tv + u)$, $v' = x'' = -u'/t$. This representation of u', v' can be written in the form (69), with

$$a_{11} = tf, \quad a_{12} = tg + t^2 f, \quad a_{21} = -f, \quad a_{22} = -g - tf. \tag{72}$$

Since (70) can be reduced to the pair of conditions (66)-(67) in the case (72), it is seen from the definition (71) of u that (γ) is a corollary of (γ bis).

Remark. Suppose that (1) is oscillatory, that is, that some (hence, according to Sturm, every) solution $x(t) \not\equiv 0$ of (1) has zeros $t = t_n$ which cluster at $t = \infty$. Then (1) *can* have a solution $x(t) \not\equiv 0$ possessing an asymptote (cf. below), but it cannot have two, linearly independent, such solutions. For, if it did, (1) would possess property (4*), hence property (4), and therefore a solution satisfying (2_1), which is impossible when (1) is oscillatory. That *one* solution ($\not\equiv 0$) possessing an asymptote *is* compatible with the oscillatory character of (1), is shown by the following example:

Draw in the (t, x)-plane a graph $x = x(t)$ in such a way that the graph of $|x(t)|$ consists of a sequence $(t_1, t_2), (t_2, t_3), \cdots$ of strictly convex arches, and that $\operatorname{sgn} x(t) = (-1)^n$ if $t_n < t < t_{n+1}$, while $x(t_n) = 0$ and $x'(t_n) \neq 0$, where $n = 1, 2, \cdots$ and $t_n \to \infty$. Then, if the function $x(t)$ possesses a continuous third derivative, a continuous function $f(t)$ is defined by placing $f(t) = -x''(t)/x(t)$ or $f(t) = f(t \pm 0)$ according as $t \neq t_n$ or $t = t_n$, since the existence of $\lim x''(t)/x(t)$, as $t \to t_n$, follows from l'Hopital's rule. With reference to this $f(t)$, the function $x(t)$ is a solution of (1), hence (1) is oscillatory. But it is clear that the successive waves of the graph of $x(t)$ can be chosen in such a way that $t_n x'(t_n) \to 0$ and $t_{n+1} - t_n < \text{const.}$ (hence $t_{n+1}/t_n < \text{Const.}$) as $n \to \infty$. Then, for reasons of convexity, $tx'(t) \to 0$ and $x(t) \to 0$ as $t \to \infty$, and so the limit (11) exists.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] G. H. Hardy, "Generalizations of a limit theorem of Mr. Mercer," *Quarterly Journal of Pure and Applied Mathematics*, vol. 43 (1912), pp. 143-150.

[2] P. Hartman, "Unrestricted solution fields of almost-separable differential equations," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 560-580.

[3] ———, "On linear second order differential equations with small coefficients," *American Journal of Mathematics*, vol. 73 (1951), pp. 955-962.

[4] O. Haupt, "Ueber Lösungen linearer Differentialgleichungen mit Asymptoten," *Mathematische Zeitschrift*, vol. 48 (1942-43), pp. 212-220; cf. pp. 289-292.

[5] E. Hilb, Article II B5 (1914) in the *Encyclopädie der mathematischen Wissenschaften*, vol. II_2.

[6] A. Wintner, "Asymptotic integrations of the adiabatic oscillator," *American Journal of Mathematics*, vol. 69 (1947), pp. 251-272.

[7] ———, "On almost free linear motions," *ibid.*, vol. 71 (1949), pp. 595-602.

[8] ———, "On linear asymptotic equilibria," *ibid.*, vol. 71 (1949), pp. 853-858.

[9] ———, "On the non-existence of conjugate points," *ibid.*, vol. 73 (1951), pp. 368-380.

LINEAR DIFFERENTIAL AND DIFFERENCE EQUATIONS WITH MONOTONE SOLUTIONS.*

By PHILIP HARTMAN and AUREL WINTNER.

---

1. A corollary of one of the principal theorems to be proved below is the following:

(i) *In the linear, homogeneous differential equation of n-th order*

$$f_0(t)D^n y + \sum_{k=1}^{n} (-1)^{k+1} f_k(t) D^{n-k} y = 0, \qquad (D = d/dt), \tag{1}$$

let the coefficient functions $f_0, \cdots, f_n$ *be continuous on* $0 < t < \infty$ *and let* f_0 *and* $f_2, \cdots, f_n$ *satisfy*

$$f_0 > 0 \text{ and } f_k \geqq 0 \text{ for } k = 2, 3, \cdots, n, \qquad (0 < t < \infty), \tag{2}$$

while $f_1 \gtreqless 0$. *Then* (1) *has at least one solution* $y = y(t)$ *which is positive and non-decreasing for* $0 < t < \infty$ *and, what is more,*

$$y > 0 \text{ and } (-1)^j D^j y \geqq 0 \text{ for every } j < n \qquad (0 < t < \infty). \tag{3}$$

The same is true if $0 < t < \infty$ *is replaced by* $0 \leqq t < \infty$.

If the second of the inequalities (2) is assumed for $k = 1$ also, then it follows from (1) and (3) that the second of the inequalities (3) remains true for $j = n$.

Assertion (i) is a generalization of a theorem of A. Kneser ([5], pp. 178-192), which assumes that $n = 2$ and that (1) is of the form $f_0 D^2 y - f_2 y = 0$, where $f_1 \equiv 0$; for the general case of $n = 2$, cf. [3]. These proofs for $n = 2$ depend on convexity arguments which are not applicable when $n > 2$. Actually, the proof of (i) below will be much simpler than the proofs of [5], [3] for the particular case $n = 2$.

It is easy to conclude that (i) has the following corollary:

(i bis) *If* $f_0 > 0$ *possesses a completely monotone derivative, and if* $f_1, \cdots, f_n$ *are completely monotone, on* $0 < t < \infty$ *(that is, if*

$$f_0 > 0, \quad (-1)^j D^{j+1} f_0 \geqq 0, \quad (-1)^j D^j f_k \geqq 0 \qquad (k = 1, \cdots, n; j = 0, 1, \cdots), \tag{4}$$

---

* Received March 31, 1953.

where $0 < t < \infty$), *then* (1) *has at least one solution* $y(t) \not\equiv 0$ *which is completely monotone for* $0 < t < \infty$:

$$(-1)^j D^j y \geqq 0 \text{ for } j = 0, 1, \cdots. \tag{5}$$

The same is true if $0 < t < \infty$ *is replaced by* $0 \leqq t < \infty$.

In other words, if each of the functions $Df_0, f_1, \cdots, f_n$ is representable as a Laplace transform

$$\int_0^\infty e^{-ts} d\mu(s), \quad \text{where } d\mu(s) \geqq 0, \tag{6}$$

for $0 < t < \infty$ or $0 \leqq t < \infty$ (Hausdorff-Bernstein), and if $f_0 > 0$, then (1) has at least one solution $y = y(t) \not\equiv 0$ which is representable in the same form on the same t-range. (i bis) was proved in the case $D^2y - f_2 y = 0$ of (1) in [8] and in the general case of $n = 2$ in [3].

(i) and (i bis) have analogues for difference equations:

(ii) *In the linear, homogeneous difference equation of n-th order*

$$f_0(m)\Delta^n y(m) + \sum_{k=1}^{n} (-1)^{k-1} f_k(m) \Delta^{n-k} y(m) = 0, \quad (m = 0, 1, \cdots), \tag{7}$$

let the coefficient sequences $f_0(m), f_1(m), \cdots, f_n(m)$ *satisfy*

$$f_0(m) > 0 \text{ and } f_k(m) \geqq 0 \text{ for } k = 2, \cdots, n, \qquad (f_1 \gtreqless 0), \tag{8}$$

where $m = 0, 1, \cdots$, *and*

$$f_0(m) - \sum_{k=1}^{n} f_k(m) > 0 \qquad (m = 0, 1, \cdots). \tag{9}$$

Then (7) *has at least one solution* $y = y(m)$ *satisfying*

$$y(m) > 0 \text{ and } (-1)^j \Delta^j y(m) \geqq 0 \text{ for } j = 1, \cdots, n-1 \qquad (m = 0, 1, \cdots). \tag{10}$$

It is understood that $\Delta y(m) = y(m+1) - y(m)$, $\Delta^2 y(m) = \Delta(\Delta y(m)) = y_{m+2} - 2y_{m+1} + y_m, \cdots$. The case $n = 2$ of (ii) was proved in [4].

It is worth making the following remarks concerning (ii):

Remark 1. If it is assumed in (ii) that $f_1(m) \geqq 0$, where $m = 0, 1, \cdots$, then (7) and (9) imply that $(-1)^n \Delta^n y(m) \geqq 0$.

Remark 2. If $n > 1$ and if it is assumed in (ii) that $f_n(m) > 0$ for an infinity of m-values, then $-\Delta y(m) > 0$ for every $m \geqq 0$.

Remark 3. If (9) is weakened to $f_0(m) - \sum_{k=1}^{n} f_k(m) \geqq 0$ and $f_0(m) - f_1(m) > 0$, then (ii) remains valid if the first inequality in (10) is correspondingly weakened to $y(m) \geqq 0$ (so that the solution in question can be of trivial type, that is, $y(m) \equiv 0$ for all sufficiently large m).

The analogue of (i bis) is as follows:

(ii bis) *If* (9) *holds and if each of the* $n+1$ *sequences* $\Delta f_0(m)$, $f_1(m), \cdots, f_n(m)$, *where* $m = 0, 1, \cdots$, *is completely montone* (*that is, if* (9) *and*

$$(11)\quad (-1)^j \Delta^{j+1} f_0(m) \geqq 0, \quad (-1)^j \Delta^j f_k(m) \geqq 0 \text{ for } k = 1, 2, \cdots, n \qquad (j, m = 0, 1, \cdots)$$

hold), *then* (7) *has a solution* $y = y(m) > 0$ *which is completely monotone, that is,*

$$(12)\qquad (-1)^j \Delta^j y(m) \geqq 0 \text{ for } j = 0, 1, \cdots \qquad (m = 0, 1, \cdots).$$

The case $n = 2$ of (ii bis) is known [4].

2. Both (i) and (i bis) will be deduced from corresponding theorems on systems of linear differential equations of first order. In order to simplify the notation, the following abbreviations will be used: If $x = (x_1, \cdots, x_n)$ is a vector, then

$$(13)\qquad x \geqq 0 \text{ means that } x_k \geqq 0 \text{ for } k = 1, \cdots, n.$$

Similarly, if $A = (a_{ik})$ is an n by n matrix, then

$$(14)\qquad A \geqq 0 \text{ means that } a_{ik} \geqq 0 \text{ for } i, k = 1, \cdots, n.$$

The analogous abbreviations $x > 0$ and $A > 0$ will also be used. If $p = (p_1, \cdots, p_n)$ is a vector, the product px will represent the vector

$$(15)\qquad px = (p_1 x_1, p_2 x_2, \cdots, p_n x_n).$$

The main theorem on differential equations to be proved is as follows:

(I) *Let* $A = A(t)$ *be an* n *by* n *matrix of continuous functions satisfying*

$$(16)\qquad A(t) \geqq 0 \qquad (0 < t < \infty).$$

Then the system of differential equations

$$(17)\qquad x' = -A(t)x \qquad (' = D = d/dt)$$

has at least one solution $x = x(t) \not\equiv 0$ *satisfying*

$$x(t) \geqq 0 \text{ and } -x'(t) \geqq 0 \qquad (0 < t < \infty). \tag{18}$$

Needless to say, $x(t)$ is defined and continuously differentiable on the *closed* range $0 \leqq t < \infty$ if $A(t)$ is defined and continuous there.

Remark 1. If $x = x(t) = (x_1(t), \cdots, x_n(t)) \not\equiv 0$ is a solution of (17) satisfying (18), then $x_k(\infty) = \lim x_k(t)$, as $t \to \infty$, exists and is non-negative for $k = 1, 2, \cdots, n$, while $-x'_k(t)$ is non-negative. Thus $-x'_k(t)$ is integrable over $1 \leqq t < \infty$. It follows therefore from (16), (17) and (18) that a necessary condition for $x_k(\infty) > 0$ to hold for some k is that the elements $a_{jk}(t)$, where $j = 1, \cdots, n$, of the k-th column of $A(t)$ be integrable over $1 \leqq t < \infty$. This necessary condition is not sufficient, as is shown by the binary system

$$x'_1 = -x_2, \qquad x'_2 = -x_1/t^2 \qquad (0 < t < \infty).$$

In fact, the general solution $x = (x_1, x_2)$ of this system is

$$x = (c_1 t^\alpha + c_2 t^\beta, \ -\alpha c_1 t^{\alpha-1} - \beta c_2 t^{\beta-1}),$$

where $\lambda = \alpha, \beta = \frac{1}{2}(1 \pm 5^{\frac{1}{2}})$ are the roots of the quadratic equation $\lambda(\lambda - 1) - 1 = 0$. But the system has (up to constant factors) only one solution $x = x(t) = (t^\beta, -\beta t^{\beta-1})$, where $\beta = \frac{1}{2}(1 - 5^{\frac{1}{2}}) < 0$, satisfying (18). For this solution, $x(\infty)$ is $(0, 0)$, although the elements of the first column, $a_{11}(t) \equiv 0$ and $a_{21} = t^{-2}$, are integrable over $1 \leqq t < \infty$.

On the other hand, *when* (16) *holds, a necessary and sufficient condition in order that* (17) *possess a solution satisfying* (18) *and* $x(\infty) > 0$ *is that every element* $a_{jk}(t)$, *where* $j, k = 1, \cdots, n$, *of* $A(t)$ *be integrable over* $1 \leqq t < \infty$. The necessity of this condition follows from the above remarks concerning $x_k(\infty) > 0$ (for some k). The sufficiency of the condition follows from a particular case of a theorem of Dunkel [1]; for a short proof of this particular case, cf. [9], pp. 262-264.

Remark 2. Let $A = A(t)$ in (I) be constant (independent of t) and let $x = x(t) \not\equiv 0$ be a solution of (17) satisfying (18). Every solution of (17), in particular, the $x(t)$ supplied by (I), is a sum of solutions of the form $x = (ct^j + O(t^{j-1}))e^{-\lambda t}$ $(\not\equiv 0)$, as $t \to \infty$, where $Ac = \lambda c$, λ is an eigenvalue of A, c a corresponding eigenvector and j a non-negative integer. Hence, the partial sum $x = \Sigma(ct^j + O(t^{j-1}))e^{-\lambda t}$ of these solutions, belonging to the λ-values with the least real part and to the greatest j associated with these λ, is real and satisfies $x \geqq 0$ for large t, since $x = x(t)$ does. In

particular, λ is real and $c \geqq 0$. Hence, $x = ce^{-\lambda t}$ is non-negative (for all t) and is a solution of (17). But (16) and (17) imply that $x' = -\lambda ce^{-\lambda t}$ is non-positive (for all t); so that $\lambda \geqq 0$.

It follows that (I) can be considered as a generalization of the algebraic theorem of Perron ([6]; cf. Frobenius [2]) which states that a non-negative (constant) matrix A possesses at least one non-negative eigenvalue λ, corresponding to which there is a non-negative eigenvector c. (There also follows Perron's result which states that if $A > 0$, then, in the last assertion, $\lambda \geqq 0$ and $c \geqq 0$ can be improved to $\lambda > 0$ and $c > 0$. In fact, $\lambda > 0$ is needed, by Remark 1, to assure that $x(t) = ce^{-\lambda t} \to 0$ as $t \to \infty$, and $c > 0$ follows from the equations $c = \lambda^{-1}Ac$, since $A > 0$ and $(0 \neq) c \geqq 0$.)

Corresponding to the preceding deduction from (I), the assertion of (i) can be considered as a generalization of a particular case of Descartes's rule for the existence of a non-positive root for a polynomial.

(I) has the following corollary:

(I bis) *Let $A = A(t)$ be a completely monotone n by n matrix and let $p(t)$ be a positive vector having a completely monotone derivative on $0 < t < \infty$; that is, let*

$$(-1)^j D^j A \geqq 0 \quad \text{for } j = 0, 1, \cdots, \tag{19}$$

and

$$p > 0 \text{ and } (-1)^j D^{j+1} p \geqq 0 \quad \text{for } j = 0, 1, \cdots, \tag{20}$$

where $0 < t < \infty$. Then

$$px' = -Ax \tag{21}$$

has at least one solution vector $x = x(t) \not\equiv 0$ which is completely monotone on $0 < t < \infty$,

$$(-1)^j D^j x \geqq 0 \quad \text{for } j = 0, 1, \cdots. \tag{22}$$

The same is true if $0 < t < \infty$ is replaced by $0 \leqq t < \infty$.

Remark. In [10], [7], linear differential equations with coefficients representable as Laplace transforms (6) *not* subject to $d\mu(s) \geqq 0$ were considered. Instead of the condition $d\mu(s) \geqq 0$ for all $s \geqq 0$, it was assumed that the contribution of small $s > 0$ to the integral (6) is small, in the sense that

$$\int_{+0} s^{-1} \, |\, d\mu(s) \,| < \infty;$$

for example, that $d\mu(s) \equiv 0$ for $0 \leqq s \leqq \epsilon$, where $\epsilon > 0$. The assertions of the theorems proved were to the effect that there exist non-trivial solutions which are (or that all solutions are) representable as Laplace transforms (6), where $d\mu \geqq 0$. The methods used were a "comparison of coefficients and a majorant method."

Theorems (i bis) and (I bis) seem to suggest that these theorems can be proved without the assumption for small $s(>0)$ on the $d\mu(s)$ occurring in the representation of the coefficients, since it might be expected that the solutions supplied by (i bis) or (I bis) furnish a suitable majorant. Since, however, the proofs of (i bis) and (I bis) do not depend on a comparison of coefficients, but on qualitative arguments leading directly to (5) and (22), this is not the case. Simple examples illustrate this negative statement; for example, in the differential equation $x' = x/t$, where $0 < t < \infty$, the coefficient function $1/t$ is the Laplace transform (6) of $\mu(s) \equiv s$, but no solution $x = \text{const.}\ t \not\equiv 0$ is a Laplace transform.

3. *Proof of* (I) *and* (I bis). Let x^0 denote a vector satisfying $x^0 > 0$; for example, let $x^0 = (1, 1, \cdots, 1)$. For a positive integer l, let $x = x^l(t)$ denote the solution of (17) satisfying the initial condition $x(t) = x^0$ when $t = l$. The assumption (16) shows that any solution $x = x(t)$ of (17) satisfies $-x'(t) \geqq 0$ on any t-interval on which $x(t) \geqq 0$. Since $x^l(l) = x^0 > 0$, it follows that $x^l(t) > 0$ for t near l, hence $x^{l\prime}(t) \leqq 0$ for t near l. Thus $x^l(t) \geqq x^l(l) > 0$ for t less than and close to l. This argument shows that $x^l(t) > 0$ and $-x^{l\prime}(t) \geqq 0$ for $0 < t \leqq l$.

If $x^l = (x^l{}_1, \cdots, x^l{}_n)$, put $a_l = \max(x^l{}_1(1), \cdots, x^l{}_n(1))$; so that $a_l > 0$. Let $z^l(t) = x^l(t)/a_l$. Then $x = z^l(t)$ is a solution of (17) and satisfies

$$z^l(t) > 0 \text{ and } -z^{l\prime}(t) \leqq 0 \text{ for } 0 < t \leqq l \tag{23}$$

and, at $t = 1$, the components of the vector $z^l(t)$ satisfy

$$0 < z^l{}_k(1) \leqq 1 \text{ for } k = 1, 2, \cdots, n; \text{ while } z^l{}_{k(l)}(1) = 1 \tag{24}$$

holds for at least one index $k = k(l)$, where $1 \leqq k(l) \leqq n$.

It is clear from (24) that there exists a sequence of integers $l_1, l_2, \cdots$ with the property that

$$z^0 = \lim_{j\to\infty} z^l(1) \text{ exists} \qquad (l = l_j); \tag{25}$$

so that

$$z^0{}_k = 1 \text{ for some } k \qquad (1 \leqq k \leqq n). \tag{26}$$

The relations (25) and (26) imply that, if $x = x(t)$ is the solution (17) determined by the initial condition $x(1) = z^0$, then $x(t) \not\equiv 0$, and that $z^l(t) \to x(t)$ as $j \to \infty$, where $l = l_j$, holds uniformly on every closed, bounded subinterval of $0 < t < \infty$. Hence, the first inequality in (18) follows from (23), while the second is a consequence of the first and of (16), (17). This proves (I).

It will be clear from the remark following (I) and from the proof of (I bis) below that it is sufficient to consider the case in which the underlying t-range is the open half-line $0 < t < \infty$. Note that, since the k-th equation in the system (21) can be divided by $p_k(t) > 0$, assertion (I) implies that (21) has at least one solution $x = x(t) \not\equiv 0$ satisfying (18), which is equivalent to the cases $j = 0$ and $j = 1$ of (22). The set of all relations (22) will be proved by induction.

The equation (21) and the product rule for differentiation give

$$-pD^{j+1}x = \sum_{i=0}^{j-1} C_{ji}(D^{j-i}p)(D^{i+1}x) + \sum_{i=0}^{j} C_{ji}(D^{j-i}A)(D^i x),$$

where the $C_{jm} = j!/m!(j-m)!$ are the binomial coefficients. Hence, if $(-1)^m D^m x \geqq 0$ for $m = 0, 1, \cdots, j$, then (20) shows that

$$(-1)^j(D^{j-i}p)(D^{i+1}x) \geqq 0$$

for $i = 0, 1, \cdots, j-1$, while (19) implies that $(-1)^j(D^{j-i}A)(D^i x) \geqq 0$ for $i = 0, 1, \cdots, j$. Thus, from $p > 0$ and the last two formula lines,

$$(-1)^{j+1}D^{j+1}x \geqq 0.$$

This proves (I bis).

4. *Proof of* (i). In the proof of (i), it can be supposed that $f_0 \equiv 1$, for otherwise (1) can be divided by $f_0 > 0$. Then, if

$$g(t) = \exp \int_1^t f_1(s)\,ds > 0, \qquad (0 < t < \infty), \tag{28}$$

(1) can be written in the form

$$g^{-1}D(gD^{n-1}y) + \sum_{k=2}^{n} (-1)^{k-1} f_k(t) D^{n-k}y = 0. \tag{29}$$

Hence, if

$$x_1 = y, x_2 = -Dy, \cdots, x_{n-1} = (-1)^{n-2}D^{n-2}y, x_n = (-1)^{n-1}gD^{n-1}y, \tag{30}$$

(29) can be written as the system of linear, first order differential equations

$$(31)\qquad x_1' = -x_2, \cdots, x_{n-2}' = -x_{n-1}, x_{n-1}' = -x_n/g, x_n' = -g\sum_{k=1}^{n-1} f_{n-k+1}x_k.$$

In view of (2) and (28), condition (16) of (I) is satisfied when (31) is identified with (17). The definitions (30) and the assertion of (I) imply that (1) has a solution $y = y(t) \not\equiv 0$ satisfying the second of the inequalities (3) for $j = 0, 1, \cdots, n-1$. In particular, $y \geqq 0$ and $Dy \leqq 0$. Hence if $y = 0$ for some value of $t = t_0 > 0$, then $y(t) \equiv 0$ for $t \geqq t_0$, and therefore for $0 < t < \infty$. Since this is a contradiction, it follows that $y(t) > 0$ for $0 < t < \infty$. This proves (3).

The transition to the case of (i) in which $0 < t < \infty$ is replaced by $0 \leqq t < \infty$ is obvious. For in the latter situation, where the coefficients of (1) are defined and continuous on $0 \leqq t < \infty$, any solution $y = y(t)$ on $0 < t < \infty$ can be defined (by continuity) at $t = 0$ so as to be a solution on $0 \leqq t < \infty$. By continuity, the last set of inequalities in (3) holds at $t = 0$ also, while the first inequality in (3) follows at $t = 0$ from the monotony of y for $t > 0$.

5. *Proof of* (i bis). Assertion (I bis) implies (i bis) if (1) is written as the system of linear, first order differential equations

$$(32)\qquad x_1' = -x_2, \cdots, x_{n-1}' = -x_n, f_0 x_n' = -\sum_{k=1}^{n} f_{n-k+1}x_k.$$

In fact, if $x_1 = y$, $x_2 = -Dy, \cdots, x_n = (-1)^{n-1}D^{n-1}y$, then (32) can be identified with (21), where p is the vector $(1, \cdots, 1, f_0)$ and every element a_{ik} of the matrix $A(t)$ is either one of the functions $f_1, f_2, \cdots, f_n$ or is identically 0 or 1. Thus (4) implies conditions (19) and (20) of (I bis), and the assertion of (I bis) is equivalent to that of (i bis), by the definition of the vector $x = (x_1, \cdots, x_n)$.

6. The theorem (I) for differential equations has an analogue for difference equations:

(II) *Let $A = A(m) = (a_{ik}(m))$ be an n by n matrix function of the non-negative integer m. Suppose that*

$$(33)\qquad A(m) \geqq 0 \qquad (0 \leqq m < \infty),$$

and that, if I is the unit matrix, the reciprocal

$$(34)\qquad (I - A(m))^{-1} \text{ exists and is } \geqq 0 \qquad (0 \leqq m < \infty).$$

Then the linear, homogeneous, vector difference equation

$$\Delta x(m) = -A(m)x(m) \tag{35}$$

has at least one solution $x = x(m)$ *satisfying*

$$x(m) \geqq 0 \text{ and } -\Delta x(m) \geqq 0 \qquad (0 \leqq m < \infty), \tag{36}$$

and

$$x(m) \neq 0 \qquad (0 \leqq m < \infty). \tag{37}$$

The relations (35) and (37) show that

$$\Delta x(m) \neq 0 \text{ if } \det A(m) \neq 0. \tag{38}$$

It will be clear from the proof of (II) that if condition (34) is replaced by the assumptions that, for some integer $l \geqq 0$,

$$(I - A(l))x = 0 \text{ has a solution, } 0 \neq x \geqq 0 \tag{34*}$$

and

$$(I - A(m))^{-1} \text{ exists and is } \geqq 0 \text{ for } 0 \leqq m < l, \tag{34**}$$

then (II) remains valid if (37) is weakened to $x(m) \not\equiv 0$. In fact, (35) will then have a solution $x = x(m)$ satisfying (36), $x(m) \neq 0$ for $m = 0, 1, \cdots, l$, and $x(m) = 0$ for $m = l+1, l+2, \cdots$.

The difference analogue of (I bis) is as follows:

(II bis) *Let* $A(m)$ *satisfy the conditions of* (II). *In addition, let* $A(m)$ *be completely monotone, and let* $p = p(m)$ *be a positive vector having a completely monotone first difference; that is, let*

$$(-1)^j \Delta^j A(m) \geqq 0 \text{ for } j = 0, 1, \cdots \qquad (0 \leqq m < \infty) \tag{39}$$

and

$$p(m) > 0 \text{ and } (-1)^j \Delta^{j+1} p(m) \geqq 0 \qquad (0 \leqq m < \infty). \tag{40}$$

Then (35) *has at least one solution* $x = x(m)$ *which satisfies* (37) *and is completely monotone,*

$$(-1)^j \Delta^j x(m) \geqq 0 \text{ for } j = 0, 1, \cdots \qquad (0 \leqq m < \infty). \tag{41}$$

7. *Proof of* (II) *and* (II bis). The equations (35) can be written in the form

$$x(m+1) = x(m) - A(m)x(m) = (I - A(m))x(m), \tag{42}$$

which, since $I - A(m)$ is non-singular, is equivalent to

$$x(m) = (I - A(m))^{-1} x(m+1). \tag{43}$$

The equations (42) and (43) show that if l is any non-negative integer and x^0 is an arbitrary vector, then (35) has a unique solution satisfying $x(l) = x^0$; in fact, (42) then determines $x(m)$ for every $m > l$, while (43) determines $x(m)$ for every $m < l$. Furthermore, (34) implies that, if $l > 0$ and if the assigned initial condition $x(l) = x^0$ is a non-negative vector, then $x(m) \geqq 0$ for $0 \leqq m \leqq l$, and so, by (35), $\Delta x(m) \leqq 0$ for $0 \leqq m \leqq l$.

The proof of (II) can now be completed along the lines of the proof of (I), as follows: Let l be a positive integer, x^0 a positive vector, say $x^0 = (1, 1, \cdots, 1)$, and $x^l = x^l(m)$ the solution of (35) satisfying $x^l(l) = x^0$. Then $x^l(m) \geqq 0$ and $-\Delta x^l(m) \geqq 0$ for $0 \leqq m \leqq l$. In particular, $x^l(0) \geqq x^l(l) > 0$, so that $a = a_l = \max(x^l{}_1(0), \cdots, x^l{}_n(0))$ is positive. Let $z^l = z^l(m) = x^l(m)/a_l$. Then $x = z^l(m)$ is a solution of (35) satisfying

$$z^l(m) > 0 \text{ and } -\Delta z^l(m) \geqq 0 \text{ and } 0 \leqq m \leqq l \tag{44}$$

and, at $m = 0$, the components of z^l satisfy $0 < z^l{}_k(0) \leqq 1$ for $k = 1, \cdots, n$, while $z^l{}_{k(l)}(0) = 1$ holds for at least one index $k = k(l)$, $0 \leqq k(l) \leqq n$.

Let $l_1, l_2, \cdots$ be an increasing sequence of integers such that

$$z^0 = \lim_{j\to\infty} z^l(0) \text{ exists} \qquad (l = l_j). \tag{45}$$

Clearly, $z^0{}_k = 1$ for some k, where $1 \leqq k \leqq n$. Let $x = x(m)$ be the solution of (35) determined by $x(0) = z^0$; in particular, $x(0) \neq 0$. If $x = x(m)$ in (42) is replaced by $x = z^l(m)$, it follows from (45) and an induction on m that

$$x(m) = \lim_{j\to\infty} z^l(m) \qquad (l = l_j) \tag{46}$$

for $m = 0, 1, \cdots$. Hence (36) follows from (44).

The remark made after (42), concerning the uniqueness of a solution of (35) satisfying a given initial condition, shows that (37) is satisfied, since $x(m) \not\equiv 0$. Hence (II) is proved.

The proof of (II bis) is similar to that of (I bis) and will be omitted; cf. [4], pp. 127-128.

8. *Proof of* (ii) *and* (ii bis). In view of the condition $f_0(m) > 0$, it can be supposed that $f_0(m) \equiv 1$. Then (9) is

$$1 - \sum_{k=1}^{n} f_k(m) > 0; \tag{47}$$

in particular, $1 - f_1(m) > 0$. Put

$$g(0) = 1 \text{ and } g(m) = g(m-1)/(1 - f_1(m-1)) > 0 \text{ for } m = 1, 2, \cdots; \tag{48}$$

so that $\Delta g(m) = g(m+1)f_1(m)$ for $m = 0, 1, \cdots$ and $\Delta(g(m)\Delta^{n-1}y(m)) = g(m+1)(\Delta^n y(m) + f_1(m)\Delta^{n-1}y(m))$. Thus (7) can be written as

$$(49) \qquad g^{-1}(m+1)\Delta(g(m)\Delta^{n-1}y(m)) + \sum_{k=2}^{n}(-1)^{k-1}f_k(m)\Delta^{n-k}y(m) = 0.$$

In terms of the vector $x = x(m) = (x_1(m), \cdots, x_n(m))$ defined by

$$(50) \qquad x_1 = y, x_2 = -\Delta y, \cdots, x_{n-1} = (-1)^{n-2}\Delta^{n-2}y, x_n = (-1)^{n-1}g\Delta^{n-1}y,$$

the equation (49) is equivalent to the system of first order difference equations

$$(51) \qquad \Delta x_1 = -x_2, \cdots, \Delta x_{n-2} = -x_{n-1}, \qquad \Delta x_{n-1} = -x_n/g,$$

$$\Delta x_n(m) = -g(m+1)\sum_{k=1}^{n-1} f_{n-k+1}(m)x_k(m).$$

If (51) is identified with (35), it is seen that (33) holds.

In order to prove the existence of $(I-A)^{-1}$, let (51) be solved for $x(m)$, if possible. To this end, put

$$(52) \qquad F_k(m) = f_k(m) + f_{k+1}(m) + \cdots + f_n(m) \geqq 0, \text{ where } k = 1, \cdots, n.$$

Multiply the k-th of the equations (51) by $g(n+1)F_{n-k+1}(m)$ or 1 according as $k = 1, \cdots, n-1$ or $k = n$, and add the resulting n equations. The result is

$$g(m+1)\sum_{k=1}^{n-1} F_{n-k+1}(m)x_k(m+1) + x_n(m+1)$$
$$= (1 - g(m+1)F_2(m)/g(m))x_n(m),$$

or since, $g(m+1)/g(m) = 1/(1-f_1(m))$,

$$(53) \qquad x_n(m) = (1-f_1(m))g(m+1)\sum_{k=1}^{n-1} F_{n-k+1}(m)x_k(m+1)/(1-F_1(m)),$$

where $1 - F_1(m) > 0$, by (47). It follows that $x_{n-1}(m), x_{n-2}(m), \cdots, x_1(m)$ can successively be expressed in terms of the components of $x(m+1)$, by using (53) and the $(n-1)$-st, the $(n-2)$-nd, $\cdots$ of the equations (51). For the first $n-1$ of the equations (51) can be written (in the reverse order) as

$$(54) \qquad x_{n-1}(m) = x_{n-1}(m+1) + x_n(m)/g(m),$$
$$x_{n-2}(m) = x_{n-2}(m+1) + x_{n-1}(m), \cdots, x_1(m) = x_1(m+1) + x_2(m).$$

Since the system (53)-(54) is equivalent to (43), it follows that $(I-A)^{-1}$ exists, in view of (47), and that $(I-A(m))^{-1} \geqq 0$, by (47) and (52).

Thus (II) is applicable to the system (51). Hence there exists a

solution $x = x(m)$ satisfying (36) and (37). It follows by (50) that (7) possesses a solution $y = y(m) \not\equiv 0$ such that (10) holds if the $>$ is replaced by $\geqq$ in the first inequality in (10). However, if $y(l) = 0$ for some $l(\geqq 0)$, then $y(m) \equiv 0$ for $m \geqq l$, since $y(m) \geqq 0$ and $-\Delta y(m) \leqq 0$. In this case, $x(m) \equiv 0$ for $m \geqq l$, by (50). Since this contradicts (37), it follows that $y(m) > 0$ for $m = 0, 1, \cdots$, and so (ii) is proved.

As to the Remark 2, following the statement of (ii), note that $\det A(m) = (-1)^n g(m+1) f_n(m)$, by (51); so that $\det A(m) = 0$ if and only if $f_n(m) = 0$. Thus, the assumption of Remark 2 and (38) show that "$\Delta x(m) \equiv 0$ for $m \geqq l$" cannot hold for any l. But if $n > 1$ and $\Delta y(m) = 0$ for $m = l$, then $\Delta y(m) = 0$, and so $\Delta x(m) \equiv 0$ for $m \geqq l$, by (50). Hence

$$-\Delta y(m) > 0 \text{ for } m = 0, 1, \cdots.$$

As to Remark 3, it is sufficient to apply the comments made above on (34*), (34**) and (II). The proof of (ii) shows that $(I - A(m))^{-1}$ exists (and is $\geqq 0$) for a given m, if and only if the inequality in (9) holds. For some m, let

$$f_0(m) - \sum_{k=1}^{n} f_k(m) = 0. \tag{55}$$

Using the notation of the proof of (ii), where $f_0(m) \equiv 1$ and where (51) is equivalent to $\Delta x(m) = -A(m)x(m)$, it is seen that $x = (1/g(m), \cdots, 1/g(m), 1)$ is a solution of $A(m)x = x$, if it is recalled that

$$g(m+1)/g(m) = 1/(1 - f_1(m))$$

and $1 - f_1(m) = F_2(m) + \cdots + F_n(m)$. Hence, if (9) is weakened to allow $\geqq$ in place of $>$, then (34*) and (34**) hold for the least $m = l \geqq 0$ satisfying (55).

The assertion (ii bis) follows from (II bis) in exactly the same way as (i bis) does from (I bis). The proof of (ii bis) will therefore be omitted.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] O. Dunkel, "Regular singular points of a system of homogeneous linear differential equations of the first order," *Proceedings of the American Academy of Arts and Sciences*, vol. 38 (1902-03), pp. 341-370.

[2] G. Frobenius, "Ueber Matrizen aus positiven Elementen," *Sitzungsberichte der Kgl. Preussischen Akademie der Wissenschaften zu Berlin*, 1908, pp. 471-476 and 1909, pp. 514-518; "Ueber Matrizen aus nicht negativen Elementen," *ibid.*, 1912, pp. 456-477.

[3] P. Hartman and A. Wintner, "On the Laplace-Fourier transcendents," *American Journal of Mathematics*, vol. 71 (1949), pp. 367-372.

[4] ——— and A. Wintner, "On linear difference equations of second order," *ibid.*, vol. 72 (1950), pp. 124-128; cf. p. 624.

[5] A. Kneser, "Untersuchung und asymptotische Darstellung der Integrale gewisser Differentialgleichungen bei grossen reellen Werthen des Arguments, I," *Journal für die reine und angewandte Mathematik*, vol. 116 (1896), pp. 178-212.

[6] O. Perron, "Grundlagen für eine Theorie des Jacobischen Kettenbruchalgorithmus," *Mathematische Annalen*, vol. 64 (1907), pp. 1-76.

[7] C. R. Putnam and A. Wintner, "Linear differential equations with almost periodic or Laplace transform coefficients," *American Journal of Mathematics*, vol. 73 (1951), pp. 792-806.

[8] A. Wintner, "On the Laplace-Fourier transcendents occurring in mathematical physics," *ibid.*, vol. 69 (1947), pp. 87-89.

[9] ———, "Asymptotic integrations of the adiabatic oscillator," *ibid.*, vol. 69 (1947), pp. 251-272.

[10] ———, "On the small divisors in integrations by Laplace transforms," *ibid.*, vol. 73 (1951), pp. 173-180.

A CONNECTION BETWEEN THE WHITEHEAD AND THE PONTRYAGIN PRODUCT.* [1]

By HANS SAMELSON.

1. Introduction. Let X be a 1-connected (i. e. arcwise and simply connected) topological space, and let Ω be the space of loops (closed paths) in X, with base point x_0 (cf. [5] for this concept). We recall some definitions and facts.

(a) If α and β are elements of the homotopy groups $\pi_{p+1}(X)$, $\pi_{q+1}(X)$, we denote, as customary, by $[\alpha, \beta]$ their Whitehead product (cf. e. g. [8]); it is an element of $\pi_{p+q+1}(X)$.

(b) The space Ω possesses a natural multiplication (composition of loops, as used in the definition of fundamental group), and this gives rise to the Pontryagin product for the (singular) homology group of Ω; if a, b are elements of $H_r(\Omega)$, $H_s(\Omega)$, then the product $a * b$ belongs to $H_{r+s}(\Omega)$ (cf. [2] for definitions and algebraic properties).

(c) There is a natural isomorphism T between $\pi_n(X)$ and $\pi_{n-1}(\Omega)$ (actually there are several such, cf. § 3); we let h denote the standard map, introduced by Hurewicz, of the homotopy groups of a space into the homology groups, and define $\tau: \pi_n(X) \to H_{n-1}(\Omega)$ by $\tau = h \circ T$. (T and τ are related to transgression, cf. [7], p. 452.) We can now state the result of the present note, with T meaning $\delta \circ p^{-1}$, cf. §§ 2, 3.

THEOREM. *If* $\alpha \in \pi_{p+1}(X)$, $\beta \in \pi_{q+1}(X)$, $p, q \geqq 1$, then

$$\tau[\alpha, \beta] = (-1)^p(\tau\alpha * \tau\beta - (-1)^{pq}\tau\beta * \tau\alpha).$$

The sign $(-1)^p$ depends of course on the choice of the map T. We remark that special cases of the formula have been known to Hurewicz, Serre, G. W. Whitehead, and also acknowledge conversations with J. C. Moore and J. Dugundji, in which the problem was raised.

We give two proofs; the first proof gives the formula as consequence of some general facts, but leaves the sign open; the second proof is very elementary and direct. In § 2 we derive a homomorphism for pairs of fiber

* Received April 29, 1953.

[1] The work on this paper was performed under OOR contract no. DA-36-034-ORD-975RD.

spaces, which seems of some interest, although the special case needed later can be given a much simpler treatment.

2. A fiber space relation. Let $p: E \to B$ be a fiber map in the sense of Serre [7], with 0-connected total space E, base B and fibers $F_x = p^{-1}(x)$. Suppose a subset E' of E is a fiber space over $p(E') = B'$ relative to $p' = p \mid E$ (p restricted to E'), with E', B' and the fibers $F'_x = p'^{-1}(x)$ again 0-connected. Let x_0 be a point of B', and put $F = F_{x_0}$, $F' = F'_{x_0}$. Clearly $F' = E' \cap F$. It is well known that the map p induces an isomorphism between the homotopy groups $\pi_n(E, F)$ and $\pi_n(B, x_0) = \pi_n(B)$ (and similarly for p') ; the proof given in [6], p. 90, applies in the present case. The symbols $\pi_1(E, F)$, $\pi_0(B)$ etc. denote sets without group structure (cf. [1], p. 167) ; p is still 1 : 1, onto in dimension 1.

We denote by T (" transgression ") the composition $\delta \circ p^{-1}$ of the isomorphism $p^{-1}: \pi_n(B) \to \pi_n(E, F)$ and the boundary map $\delta: \pi_n(E, F) \to \pi_{n-1}(F)$; we have the analogous T for E'. We now construct a similar map for the relative groups. The spaces (E, E', F) form a triad (in the sense of [1]). We consider the following diagram (it applies to an arbitrary triad) :

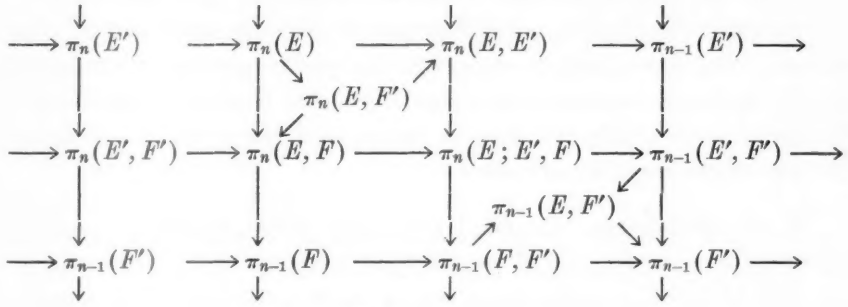

DIAGRAM 1.

All maps are the standard maps; commutativity clearly holds everywhere except that for $n > 2$ the two maps from $\pi_n(E; E', F)$ to $\pi_{n-1}(E, F')$ are negatives of each other (cf. Lemma 3. 5. 5 of [1], p. 177 for the proof). On the other hand, the map p maps the triad $(E; E', F)$ into the triad $(B; B', x_0)$ and therefore it maps the middle row of Diagram 1 into the homotopy sequence of (B, B') (i. e. the commutativity relations hold) ; in view of the isomorphisms noted above and the " five lemma " ([2], p. 16) this is actually an isomorphism of the two sequences ($\pi_2(E; E', F)$ has to be treated separately). We compose the inverse p^{-1} of this isomorphism with the maps δ from the middle row to the bottom row of Diagram 1, and put $T = \delta \circ p^{-1}$ (" transgression "). The map T can be made into a homo-

morphism of the respective sequences, i. e. commutativity can be introduced, by changing the sign in every other dimension.

PROPOSITION 1. *There exists a homomorphism T_0 of the homotopy sequence of (B, B') into that of (F, F') of degree -1, i. e. which lowers dimension by one; the maps of $\pi_n(B')$ into $\pi_{n-1}(F')$, resp. of $\pi_n(B)$ into $\pi_{n-1}(F)$ are given by $(-1)^n\delta \circ p'^{-1}$, resp. $(-1)^n\delta \circ p^{-1}$. (We recall that $\pi_1(F, F')$ has no group structure; $\pi_0(F')$ and $\pi_0(F)$ reduce to the neutral element.)*

COROLLARY. *If E' and E are contractible, then the homotopy sequences of (B, B') and (F, F') are isomorphic with a shift of 1 in dimension (for dimension 1 this reduces to the fact that $\pi_1(B')$, $\pi_1(B)$, $\pi_0(F')$, $\pi_0(F)$ are all trivial).*

Proof of the Corollary. The maps δ are now isomorphisms, as seen from Diagram 1.

Proposition 1 applies, appropriately interpreted, to the slightly more general case of a fiber map f of a fiber space (E', B', p') into another (E, B, p) (i. e. a pair of maps $f_e: E' \to E$, $f_b: B' \to B$, such that $f_b \circ p' = p \circ f_e$). The relative groups $\pi_n(B, B')$, $\pi_n(F, F')$ have to be understood as the relative groups of the corresponding maps, i. e. the relative groups of the mapping cylinder modulo the mapped space; similarly for the triad groups $\pi_n(E; E', F)$. The proof is essentially the same as before, modified by the introduction of the appropriate mapping cylinders.

3. A special case. Let (X, Y) be a pair, with both spaces 1-connected; take a point x_0 in Y. Let E' be the space of paths in Y, which end at x_0 (cf. [1], [7]); it is a fiber space over Y, with projection p' (p' maps each path on its initial point), and fiber Ω_Y (space of loops in Y, based at x_0); let E, p, Ω_X be the corresponding objects for X. Then E' is contained naturally in E, and p' is the restriction of p to E', so that we are in the situation of § 2, with Ω_Y, Ω_X playing the roles of F', F. Since E and E' are contractible ([7], p. 471), we can apply the corollary of § 2, and obtain an isomorphism T_0 of degree -1, of the homotopy sequence of (X, Y) onto that of (Ω_X, Ω_Y).

Actually this can be established in a much simpler way: Let i and n be integers with $1 < i \leqq n$. If f is a map from the n-cube I^n (product of n copies of the interval $I = [0, 1]$) to X, with the boundary $\dot{I}^n$ going into x_0, we define $T_i f$ as the map from I^{n-1} to Ω_X given by

$$T_i f(x_1, \cdots, x_{n-1})(t) = f(x_1, \cdots, x_{i-1}, t, x_i, \cdots, x_{n-1})$$

(the definition makes sense, if only the faces $x_i = 0$ and $x_i = 1$ of I^n go into x_0). Clearly the operator T_i induces an isomorphism, also called T_i, between $\pi_n(X) = \pi_n(X, x_0)$ and $\pi_{n-1}(\Omega_X)$ (with base point e_0, where e_0 is the degenerate path ($e_0(t) = x_0$ for $0 \leqq t \leqq 1$)); one could even allow $i = 1$, replacing addition in $\pi_{n-1}(\Omega_X)$ by the multiplication in Ω_X mentioned in § 1, (b). For $i < n$ the operator T_i defines an isomorphism of the relative groups $\pi_n(X, Y)$ and $\pi_{n-1}(\Omega_X, \Omega_Y)$ (this is, for $i = n - 1$, a special case of a remark in [4], p. 493), and commutes with the boundary operator; in particular T_2 maps the homotopy sequence of (X, Y) isomorphically onto that of (Ω_X, Ω_Y), with degree -1. (For the group $\pi_2(X, Y)$ one has to replace T_2 by $-T_1$; it is then mapped 1:1 onto $\pi_1(\Omega_X, \Omega_Y)$.) One verifies that for $1 \leqq i, j \leqq n$ the relations $T_i = (-1)^{i+j} T_j$ hold; this follows from the fact that the interchange of two axes is an orientation reversing homeomorphism of I^n. The operation T_n, applied to $\pi_n(X)$, coincides with the map $\delta \circ p^{-1}$ considered in § 2: If f is a map of $(I^n, \dot{I}^n)$ into (X, x_0), we define a map f' of I^n into E by $f'(x_1, \cdots, x_n)(t) = f(x_1, \cdots, x_{n-1}, t + x_n(1 - t))$. Clearly $p \circ f' = f$, and if the point $(x_1, \cdots, x_n)$ belongs to $\dot{I}^n - I^{n-1}$, then $f'(x_1, \cdots, x_n) = e_0$; so that f' defines an element of $\pi_n(E, \Omega_X)$, projecting into the element of $\pi_n(X)$ defined by f. Applying δ, i. e. putting $x_n = 0$, we obtain a map f'' of $(I^{n-1}, \dot{I}^{n-1})$ into (Ω_X, e_0), given by $f''(x_1, \cdots, x_{n-1})(t) = f(x_1, \cdots, x_{n-1}, t)$, which is exactly $T_n f$.

4. An application. We apply § 3 to the case where X is the cartesian product of two (oriented) spheres S^{p+1}, S^{q+1} of dimensions $p+1, q+1$ ($p, q \geqq 1$), and where Y is the "union" $S^{p+1} \vee S^{q+1}$, i. e. $S^{p+1} \times z_0 \cup y_0 \times S^{q+1}$, with $y_0 \varepsilon S^{p+1}$, $z_0 \varepsilon S^{q+1}$. We recall some known facts [8]. The Künneth formula (for a proof cf. [2]), applied to the well-known homology groups of spheres, shows that the first non-vanishing relative homology group of X mod Y occurs in dimension $p + q + 2$, and that this group is infinite cyclic; by the relative Hurewicz theorem [5] the same situation holds in homotopy. Further, $\pi_n(Y)$ maps onto $\pi_n(X)$ in all dimensions, so that, by exactness, δ maps $\pi_{p+q+2}(X, Y)$ isomorphically into $\pi_{p+q+1}(Y)$; i. e. the kernel of the map $\pi_{p+q+1}(Y) \to \pi_{p+q+1}(X)$ is infinite cyclic; if α_0, resp. β_0 are the elements of $\pi_{p+1}(S^{p+1})$, resp. $\pi_{q+1}(S^{q+1})$, defined by the identity maps, and α_1, β_1 are their images, under inclusion, in Y, the Whitehead product $[\alpha_1, \beta_1]$ is a generator of this kernel (this is implicit in the definition of [,], as formulated in [8], p. 201). We now apply $T_0 (= T_2)$; it follows that the first non-vanishing relative homotopy and homology group of (Ω_X, Ω_Y) occur in dimension $p + q + 1$, and that they are isomorphic and infinite cyclic.

(For p or $q = 1$ one has to use here the fact, that Ω_X, as H-space, is simple, [7], p. 479.) We have the diagram

$$\begin{array}{ccccccc} \longrightarrow & \pi_{p+q+1}(\Omega_X, \Omega_Y) & \xrightarrow{\delta} & \pi_{p+q}(\Omega_Y) & \longrightarrow & \pi_{p+q}(\Omega_X) & \longrightarrow \\ & \uparrow\downarrow & & \downarrow h & & \downarrow & \\ \longrightarrow & H_{p+q+1}(\Omega_X, \Omega_Y) & \xrightarrow{\delta} & H_{p+q}(\Omega_Y) & \xrightarrow{i} & H_{p+q}(\Omega_X) & \longrightarrow \end{array}$$

DIAGRAM 2.

The image of δ in the upper line is the infinite cyclic group generated by $T[\alpha_1, \beta_1]$. We now determine the image of δ in the lower line, i. e. the kernel of i.

5. The Pontryagin ring. We recall the concept of Pontryagin ring, defined for the space Ω of loops in any topological space: the natural composition or multiplication of loops (denoted by $\cdot$) can be considered as a map γ from $\Omega \times \Omega$ to Ω, and induces a multiplication (the Pontryagin product) in the homology group of Ω (cf. [2] for details). The Pontryagin product $a * b$ of two homology classes of Ω is the image $\gamma(a \otimes b)$ of the element $a \otimes b$ of $H(\Omega \times \Omega)$.

The Pontryagin rings of the loop spaces Ω_1, Ω_2 of S^{p+1}, S^{q+1}, are polynomial rings in the variables $\tau\alpha_0$, $\tau\beta_0$ (of dimensions p, q) (cf. [2], II. 1. 3 or [9], p. 215). The space Ω_X of our present X is clearly the cartesian product of Ω_1 and Ω_2; this implies that $H(\Omega_X) = H(\Omega_1) \otimes H(\Omega_2)$, since $H(\Omega_1)$ and $H(\Omega_2)$ are free groups. Moreover, Ω_X is the direct product of Ω_1 and Ω_2 with respect to the multiplication γ. It follows easily that the Pontryagin ring of Ω_X is the (skew) tensor product of those of Ω_1 and Ω_2, i. e. that multiplication satisfies the rule

$$(a \otimes b) * (c \otimes d) = (-1)^{rs}(a * c) \otimes (b * d),$$

with $r = \dim b$, $s = \dim c$. (One shows, that for the map $\lambda: \Omega_1 \times \Omega_2 \to \Omega_2 \times \Omega_1$, given by $\lambda(x, y) = (y, x)$, one has $\lambda(c \otimes b) = (-1)^{rs} b \otimes c$; the desired property of $*$ follows then from the commutative diagram

$$\begin{array}{ccc} (\Omega_1 \times \Omega_2) \times (\Omega_1 \times \Omega_2) & \xrightarrow{\gamma} & \Omega_1 \times \Omega_2 \\ \uparrow\downarrow & & \uparrow\downarrow \\ (\Omega_1 \times \Omega_1) \times (\Omega_2 \times \Omega_2) & \xrightarrow{\gamma \times \gamma} & \Omega_1 \times \Omega_2.) \end{array}$$

DIAGRAM 3.

We can identify $\tau\alpha_0$, $\tau\beta_0$ with the elements $\tau\alpha_0 \otimes e_0$, $e_0 \otimes \tau\beta_0$ in $H_*(\Omega_X)$

$= H_*(\Omega_1) \otimes H_*(\Omega_2)$, so that $H_*(\Omega_X)$ is the ring (with unit e_0) generated by $\tau\alpha_0$, $\tau\beta_0$, subject to the relation

$$\tau\alpha_0 * \tau\beta_0 = (-1)^{pq}\tau\beta_0 * \tau\alpha_0 \ (= \tau\alpha_0 \otimes \tau\beta_0).$$

On the other hand, it has been shown in [2], III. 1. B., that the Pontryagin ring $H_*(\Omega_Y)$ of the present Y is the free associative algebra (with unit e_0) generated by the two elements $\tau\alpha_1$, $\tau\beta_1$, the (spherical) homology elements determined by $T\alpha_1$, $T\beta_1$. The inclusion $i\colon \Omega_Y \subset \Omega_X$ induces a homomorphism of the Pontryagin rings, since it is homomorphic with respect to γ, and clearly the elements $\tau\alpha_1$, $\tau\beta_1$ are mapped into $\tau\alpha_0$, $\tau\beta_0$. It is algebraically obvious that in dimension $p+q$ the kernel of i is the infinite cyclic group generated by $\tau\alpha_1 * \tau\beta_1 - (-1)^{pq}\tau\beta_1 * \tau\alpha_1$. Our main result, for the space Y under consideration, now follows immediately from Diagram 2: Since the groups on the left are isomorphic, h must map the generator $T[\alpha_1, \beta_1]$ of the infinite cyclic group $\delta\pi_{p+q+1}(\Omega_X, \Omega_Y)$ onto $\pm$ the generator

$$\tau\alpha_1 * \tau\beta_1 - (-1)^{pq}\tau\beta_1 * \tau\alpha_1$$

of the infinite cyclic group $\delta H_{p+q+1}(\Omega_X, \Omega_Y)$.

6. The general case. Let now X stand for an arbitrary 1-connected space, and let α, β be elements of $\pi_{p+1}(X), \pi_{q+1}(X)$. We represent α and β by maps of S^{p+1} and S^{q+1} into X, and construct the obvious map f of $S^{p+1} \vee S^{q+1}$ into X. Then the elements α_1, β_1 of § 4 map into α, β under f. There is an induced map f' of the loop space of $S^{p+1} \vee S^{q+1}$ into that of X; f and f' are homomorphic with respect to $[\ ,\ ]$ and $*$; f' commutes with h; and we have $f' \circ T = T \circ f$. The result for X now follows by applying f' to the result for $S^{p+1} \vee S^{q+1}$, which holds by § 5.

7. Second proof. Our second proof is based on an interpretation of the Whitehead product, which goes back to Hurewicz and G. W. Whitehead; we present a derivation of this interpretation. X is again a 1-connected space, x_0 a point in it. We recall that in Ω_X an inversion σ, $(x \to x^{-1})$ is defined (by replacing the parameter t of any loop by $1-t$); clearly $\sigma^2 = 1$; the map $x \to x \cdot x^{-1}$, i. e. $\gamma \circ 1 \times \sigma \circ \Delta$, where Δ is the diagonal map of Ω_X into $\Omega_X \times \Omega_X$, is null-homotopic ([9], p. 210).

Let f and g be maps of $(I^{p+1}, \dot{I}^{p+1})$ and $(I^{q+1}, \dot{I}^{q+1})$ into (X, x_0), representing elements α and β of the respective homotopy groups. Then $[\alpha, \beta]$ is given by the map k of

$$S = S^{p+q+1} = \dot{I}^{p+q+2} = (I^{p+1} \times I^{q+1})^{\cdot} = \dot{I}^{p+1} \times I^{q+1} \cup I^{p+1} \times \dot{I}^{q+1},$$

defined by $k(x, y) = f(x)$, if $y \in \dot{I}^{q+1}$, $= g(y)$, if $x \in \dot{I}^{p+1}$; the base point on

S is $\omega = (0, \cdots, 0)$. We write $I^{p+1} = I^p \times I$, $I^{q+1} = I^q \times I$, and $I^{p+q+2} = I^p \times I \times I^q \times I$. Let K be the subset of S, given by

$$I^p \times I \times \dot{I}^q \times I \cup \dot{I}^p \times I \times I^q \times I \cup I^p \times 0 \times I^q \times 0;$$

by collapsing the two factors I in the first two sets in this union one can contract K over itself into $I^p \times 0 \times I^q \times 0$, and then into ω, with ω stationary. We now construct a map ϕ of $E = I^{p+q+1} = I^p \times I^q \times I$ into S as follows: for any $x \in I^p$, $y \in I^q$ the interval $x \times y \times I$ is divided into 6 parts by the t-values $0, \frac{1}{4}, \frac{3}{8}, \frac{1}{2}, \frac{5}{8}, \frac{3}{4}, 1$ and mapped in the obvious piecewise linear fashion on the closed polygon in S with successive vertices

$$\omega, (x, 0, y, 0), (x, 1, y, 0), (x, 1, y, 1), (x, 0, y, 1)\,(x, 0, y, 0), \omega.$$

If either $x \in \dot{I}^p$ or $y \in \dot{I}^q$ or $t = 0$ or $t = 1$, then $\phi(x, y, t) \in K$; so that $\phi(\dot{E}) \subset K$. For $x \in I^p - \dot{I}^p$, $y \in I^q - \dot{I}^q$, $\frac{1}{4} < t < \frac{3}{8}$ one verifies that (x, y, t) is the only point of E mapped onto $(x, 8t - 2, y, 0)$ by ϕ, that ϕ is locally $1:1$ in the neighborhood of (x, y, t) and that the local degree, using the natural orientations of E and S, is $(-1)^p$. It follows that the generator ϵ of $\pi_{p+q+1}(E, \dot{E})$, represented by the identity map of E, is mapped by ϕ into $(-1)^p\eta$, where η is the generator of $\pi_{p+q+1}(S, K)$ represented by the identity map of S.

Put $F = I^p \times I^q$ so that $E = F \times I$. Let s_1, s_2 be the customary maps of I^p, I^q onto S^p, S^q by collapsing the boundary to a point; let $s = s_1 \times s_2$ the induced map of $(F, \dot{F})$ onto $(S^p \times S^q, S^p \vee S^q)$. Let $T = T_{p+q+1}$, as defined in § 3, so that $T(k \circ \phi)(x, y)(t) = k \circ \phi(x, y, t)$. One verifies that $T(k \circ \phi)$ can be factored in the form $c \circ s$ (if $x \in \dot{I}^p$, then $T(k \circ \phi)(x, y)$ depends on y only, if $y \in \dot{I}^q$, then $T(k \circ \phi)(x, y)$ depends on x only); here c is defined as follows: the maps $T_{p+1}f$, $T_{q+1}g$ can be factored into $f' \circ s_1$, $g' \circ s_2$, with f', g' mapping S^p, S^q into Ω_X; then $c(x, y)$ is

$$(e \cdot (f'(x) \cdot g'(y)) \cdot ((f'(x)^{-1} \cdot g'(y)^{-1}) \cdot e).$$

Clearly c is homotopic to the map d, defined by

$$d(x, y) = (f'(x) \cdot g'(y)) \cdot (f'(x)^{-1} \cdot g'(y)^{-1}).$$

We now contract K over itself to ω, as indicated above, and extend this to a deformation ψ_t of S, with $\psi_0 =$ identity. One verifies that for each t the map $T(k \circ \psi_t \circ \phi)$ can be factored in the form $c_t \circ s$; the c_t consequently form a homotopy of $c = c_0$. Clearly $\psi_1 \circ \phi$ maps ϵ into $(-1)^p\eta$, just as ϕ did. But $\psi_1 \circ \phi(\dot{E}) = \omega$, so that $\psi_1 \circ \phi$ maps $(E, \dot{E})$ with degree $(-1)^p$ into (S, ω); it follows that $k \circ \psi_1 \circ \phi$ represents $(-1)^p[\alpha, \beta]$. By definition then the map $T(k \circ \psi_1 \circ \phi)$ maps the natural generator of $H_{p+q}(F, \dot{F})$ into $(-1)^p\tau[\alpha, \beta]$ (we identify here $H_{p+q}(\Omega_X)$ and $H_{p+q}(\Omega_X, e_0)$). Since this

map factors into $c_1 \circ s$, it follows that c_1, and therefore also c and d, map the natural generator of $H_{p+q}(S^p \times S^q)$ into $(-1)^p\tau[\alpha, \beta]$; we have used the fact that s maps $H_r(F, F)$ isomorphically onto $H_r(S^p \times S^q, S^p \vee S^q)$ (cf. [3], p. 266), that the map from $H_{p+q}(S^p \times S^q)$ to the relative group in the homology sequence of $(S^p \times S^q, S^p \vee S^q)$ is an isomorphism onto, and that c_1 maps $S^p \vee S^q$ into the point e_0. Our problem is now reduced to the discussion of the homology type of d.

8. The homology type of d. Let δ_1, δ_2 be the diagonal maps of S^p into $S^p \times S^p$, resp. S^q into $S^q \times S^q$; let λ be the permutation map of $S^p \times S^q$ onto $S^q \times S^p$; σ and γ are still inversion and multiplication in Ω_X. Then d can be written as the composition of the following maps: $\delta_1 \times \delta_2$, $1 \times \lambda \times 1$, $f' \times g' \times f' \times g'$, $1 \times 1 \times \sigma \times \sigma$, $\gamma \times \gamma$, γ. Let a, b denote the natural generators of $H_p(S^p)$, $H_q(S^q)$, so that by definition $f'(a) = \tau\alpha$, $f'(b) = \tau\beta$; we have to determine $d(a \otimes b)$. We use e as a generic symbol for the homology class of a point. It is well known that

$$\delta_1(a) = a \otimes e + e \otimes a, \text{ and } \delta_2(b) = b \otimes e + e \otimes b,$$

so that

$$\begin{aligned}\delta_1 \times \delta_2(a \otimes b) \\ = a \otimes e \otimes b \otimes e + a \otimes e \otimes e \otimes b + e \otimes a \otimes b \otimes e + e \otimes a \otimes e \otimes b.\end{aligned}$$

Applying a remark of § 5, we have $\lambda(a \otimes b) = (-1)^{pq} b \otimes a$, $\lambda(a \otimes e) = e \otimes a$, $\lambda(e \otimes b) = b \otimes e$. As for σ, we note that, in consequence of a remark at the beginning of § 7, the map $\gamma \circ 1 \times \sigma \circ f' \times f' \circ \delta_1 = \gamma \circ 1 \times \sigma \circ \Delta \circ f'$ is homotopic to 0. Using the equation $\delta_1(a) = a \otimes e + e \otimes a$ and the fact that e is unit element in the Pontryagin ring of Ω_X, one finds $f'(a) + \sigma \circ f'(a) = 0$. If one now applies the factors of the map d in succession to $a \otimes b$ and makes use of the relations just stated, one obtains the result

$$d(a \otimes b) = \tau\alpha * \tau\beta - (-1)^{pq}\tau\beta * \tau\alpha.$$

As remarked at the end of § 7, we have $d(a \otimes b) = (-1)^p\tau[\alpha, \beta]$, and our main result is established.

9. Remarks.

1. The relation $[\alpha, \beta] = (-1)^{(p+1)(q+1)}[\beta, \alpha]$ is known to hold [8]; an easy computation shows that the right side of our main relation, with the sign as determined, satisfies the same relation.

2. If X is an n-sphere, and $\alpha = \beta = \iota_n$ are the elements of $\pi_n(S^n)$ determined by the identity map, we have

$$\tau[\iota_n, \iota_n] = \begin{cases} -2\tau\iota_n * \tau\iota_n, & n \text{ even} \\ 0, & n \text{ odd.} \end{cases}$$

This expresses the fact that the Hopf invariant of $[\iota_n, \iota_n]$ is ± 2 or 0, depending on the parity of n.

3. If $\gamma \in \pi_{r+1}(X)$, with α and β as before, then one verifies that the τ-image of

$$[\alpha, \beta, \gamma] = (-1)^{(p+1)r}[\alpha, [\beta, \gamma]] + (-1)^{(q+1)p}[\beta, [\gamma, \alpha]] + (-1)^{(r+1)q}[\gamma, [\alpha, \beta]]$$

vanishes. It seems a reasonable conjecture that the homotopy element $[\alpha, \beta, \gamma]$ itself vanishes; the validity of this Jacobi identity remains an open question.

4. The computation of § 8 is valid in any space, which possesses a continuous multiplication with a homotopy-unit and an inversion (such as γ, e_0, σ for Ω_X).

THE INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

[1] A. L. Blakers and W. S. Massey, "The homotopy groups of a triad I," *Annals of Mathematics*, vol. 53 (1951), pp. 161-205.

[2] R. Bott and H. Samelson, "On the Pontryagin product in spaces of paths," to appear in *Comm. Math. Helv.*

[3] S. Eilenberg and N. E. Steenrod, *Foundations of algebraic topology*, Princeton, 1952.

[4] R. H. Fox, "Homotopy and torus homotopy groups," *Annals of Mathematics*, vol. 49 (1948), pp. 471-510.

[5] S. T. Hu, "An exposition of the relative homotopy theory," *Duke Mathematical Journal*, vol. 14 (1947), pp. 991-1033.

[6] N. E. Steenrod, *The topology of fiber bundles*, Princeton, 1951.

[7] J. P. Serre, "Homologie singulière des espaces fibrés, applications," *Annals of Mathematics*, vol. 54 (1951), pp. 425-505.

[8] G. W. Whitehead, "A generalization of the Hopf invariant," *Annals of Mathematics*, vol. 51 (1950), pp. 192-237.

[9] ———, "On the Freudenthal theorems," *Annals of Mathematics*, vol. 57 (1953), pp. 209-228.

GALOIS THEORY OF DIFFERENTIAL FIELDS.*

By E. R. KOLCHIN.[1]

TABLE OF CONTENTS.

* Received April 21, 1953. This paper was an address delivered before the New York meeting of the American Mathematical Society on April 26, 1952, by the invitation of the Committee to Select Hour Speakers for Eastern Sectional Meetings; it was originally submitted to the Bulletin of the American Mathematical Society on October 29, 1952, but it could not be published among the invited addresses in the Bulletin because of its length.

[1] Part of the research on which this paper is based was done in connection with a contract with the Office of Naval Research.

Introduction.

1. Differential fields and meaning of normality. The study of algebraic equations has led to the concept of field, and thence to the beginnings of algebraic geometry and to Galois theory. In much the same way, the study of algebraic differential equations has, in modern times, led to the concept of *differential field*, and thence, in the work of the late J. F. Ritt, to the extensive theory of differential algebra, which in its elementary parts bears considerable analogy to the elementary parts of algebraic geometry (see Ritt [8]). A differential field is a commutative field, in the usual sense, together with a finite family of operators $\delta_1, \cdots, \delta_m$ each of which maps the field into itself as a derivation and which commute in pairs; a differential field is said to be *ordinary* or *partial* according as the number m equals or exceeds 1.[2] *In the present paper the expression "differential field" always stands for "differential field of characteristic* 0." The purpose of the present paper is to develop a Galois theory for such differential fields.

A main problem in initiating such a theory is to find a suitable definition of normal extension of a differential field. Now, two special cases of a Galois theory already exist, and it is natural to look to these examples for hints, and to require that any general theory developed generalize these two cases. One of these cases is the Galois theory of differential field extensions of finite degree, which is the classical Galois theory, since a relative field isomorphism of such an extension is automatically a differential field isomorphism. The other case is the Picard-Vessiot theory (see Kolchin [3] and [6]; a certain

[2] By permitting m to be 0 it is possible to subsume the concept of field under that of differential field; we shall not pursue this possibility further here, and when we refer to a differential field it will always be understood that $m \geq 1$.

familiarity of the reader with the contents of these two papers will be assumed).

In the classical Galois theory an algebraic field extension of characteristic 0 is normal if the field of invariants of the group of all automorphisms of the extension over the ground-field is the ground-field itself. If an extension has this property it follows that it has the same property when considered as an extension of any intermediate field; indeed, the fundamental theorem of Galois theory could not hold were this not the case. When we turn to differential fields, however, the state of affairs is different. If $\mathfrak{F}$ is a differential field and $\mathfrak{G}$ is an extension of $\mathfrak{F}$ such that every invariant of the group of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ belongs to $\mathfrak{F}$, and if $\mathfrak{F}_1$ is a differential field between $\mathfrak{F}$ and $\mathfrak{G}$, it does not follow, even if $\mathfrak{G}$ is finitely generated and differentially algebraic over $\mathfrak{F}$, that every invariant of the group of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}_1$ belongs to $\mathfrak{F}_1$ (see the example in footnote 7). Accordingly, we define $\mathfrak{G}$ to be *weakly normal* over $\mathfrak{F}$ if the invariants of the group of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ all belong to $\mathfrak{F}$, and $\mathfrak{G}$ to be *normal* over $\mathfrak{F}$ if $\mathfrak{G}$ is weakly normal over every differential field between $\mathfrak{F}$ and $\mathfrak{G}$. The latter is the same definition as given in Kolchin [3], § 16; in that paper it was shown, and indeed it is obvious, that when $\mathfrak{G}$ is normal over $\mathfrak{F}$ in this sense then there is a one-to-one Galois correspondence between the set of all differential fields intermediate to $\mathfrak{F}$ and $\mathfrak{G}$ and a *certain* set of subgroups of the group $\mathfrak{G}$ of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$. If the subgroup corresponding to an intermediate differential field is normal then the intermediate differential field is a normal extension of $\mathfrak{F}$ (but not conversely!). Aside from the fact that in this definition of normality we demand what is essentially the conclusion of the theorem we wish to prove, there remain two blemishes, one of which we can remove, the other of which we can not. The first blemish is that, when the subgroup $\mathfrak{G}(\mathfrak{F}_1)$ of $\mathfrak{G}$ corresponding to $\mathfrak{F}_1$ is normal, so that $\mathfrak{F}_1$ is a normal extension of $\mathfrak{F}$, the factor group $\mathfrak{G}/\mathfrak{G}(\mathfrak{F}_1)$ need not be isomorphic with the group of all automorphisms of $\mathfrak{F}_1$ over $\mathfrak{F}$. This situation is remedied by defining a set of isomorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ to be *abundant* if, for every intermediate differential field $\mathfrak{F}_1$ and every element α of $\mathfrak{G}$ not in $\mathfrak{F}_1$, there exists an isomorphism σ in the set which leaves every element of $\mathfrak{F}_1$ invariant but which does not leave α invariant; clearly $\mathfrak{G}$ is normal if and only if the group of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ is abundant. Furthermore, if $\mathfrak{G}$ is normal over $\mathfrak{F}$, and $\mathfrak{G}$ is *any* abundant group of automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ (not necessarily the full automorphism group), the above mentioned results continue to hold and, when $\mathfrak{G}(\mathfrak{F}_1)$ is a normal subgroup of $\mathfrak{G}$, so that $\mathfrak{F}_1$ is a normal extension of $\mathfrak{F}$, then $\mathfrak{G}/\mathfrak{G}(\mathfrak{F}_1)$ is isomorphic to an abundant group of automorphisms of $\mathfrak{F}_1$ over $\mathfrak{F}$. The

second and more serious blemish is that we have no characterization of those "certain" subgroups which correspond to the intermediate differential fields. To avoid this defect we seek a more stringent definition.

In the classical Galois theory a normal extension is characterized also by the property that every relative isomorphism of the extension into any overfield of the extension is actually an automorphism. It would be unreasonable to demand the analogous property for differential fields, as this would exclude even the Picard-Vessiot extensions; indeed it can be shown that the extension would then be a normal algebraic extension in the classical sense. However, a hint of how to proceed is contained in the Picard-Vessiot theory. Let $\mathfrak{G}$ be a Picard-Vessiot extension of $\mathfrak{F}$, presupposing thereby that $\mathfrak{F}$ and $\mathfrak{G}$ are subject to the restrictions that $\mathfrak{F}$ and $\mathfrak{G}$ have the same field of constants $\mathfrak{C}$, that $\mathfrak{C}$ is algebraically closed, and that $\mathfrak{G}$ is finitely generated and of finite transcendence degree over $\mathfrak{F}$; it is easy to verify that $\mathfrak{F}$ and $\mathfrak{G}$ have the property that if σ is any isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$ into an extension of $\mathfrak{G}$ and if $\mathfrak{C}_\sigma$ denotes the field of constants of the compositum $\mathfrak{G}\langle\sigma\mathfrak{G}\rangle$ then

$$\mathfrak{G}\langle\sigma\mathfrak{G}\rangle = \mathfrak{G}\langle\mathfrak{C}_\sigma\rangle = (\sigma\mathfrak{G})\langle\mathfrak{C}_\sigma\rangle. \tag{1}$$

This is the property which we use, in the general case subject to the above restrictions, for our definition, which may be formulated in the following manner. We define an isomorphism σ of $\mathfrak{G}$ into an extension of $\mathfrak{G}$ to be *strong* if (1) holds; obviously every automorphism is a strong isomorphism. We then say, when $\mathfrak{F}$ and $\mathfrak{G}$ are subject to the above restrictions, that $\mathfrak{G}$ is *strongly normal* over $\mathfrak{F}$ if every isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$ is strong.[3] As indicated in the following summary, it is this type of normality which appears to be the fruitful one.

2. Summary. Chapter I contains various results from elementary differential algebra which are used in the succeeding chapters. Several of these extend to partial differential fields results which are already known in the ordinary case. One theorem proved asserts the existence, for any differential field, of a suitably defined *universal extension*; roughly speaking, a given extension is universal if it is so big that all elements of all extensions we ever have occasion to introduce may be taken in the given extension. The use of a universal extension, which follows the now well-known procedure of

[3] By pursuing further the possibility mentioned in footnote [2] and suitably formulating the definition of constant, we could make the classical concept of normal algebraic extension of finite degree of a field of characteristic 0 a special case of concept of strongly normal extension of a differential field.

modern algebraic geometry (see Weil [9]), makes it possible to avoid certain logical difficulties connected with phrases like "the set of all extensions" (of a given differential field).

Chapter II contains a detailed study of strong isomorphisms of an extension $\mathcal{G}$ of a differential field $\mathcal{F}$, subject to the restrictions above. It is shown that it is possible, in a natural way, to introduce a multiplication in the set $\mathfrak{G}^*$ of all strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$, with respect to which $\mathfrak{G}^*$ becomes a group; the group $\mathfrak{G}$ of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ is then a subgroup of $\mathfrak{G}^*$. The concept of specialization is defined for an isomorphims of $\mathcal{G}$ into an extension of $\mathcal{G}$ (or more generally, for a family of such isomorphisms): σ' is called a *specialization* of σ if the family of elements $(\sigma'\alpha)_{\alpha \in \mathcal{G}}$ is a specialization over $\mathcal{G}$ of the family $(\sigma\alpha)_{\alpha \in \mathcal{G}}$. If σ is a strong isomorphism of $\mathcal{G}$ over $\mathcal{F}$ so is every specialization of σ. Most of the important facts concerning specializations of strong isomorphisms follow from Proposition 9 of Chapter II, which asserts that if $\sigma_1, \cdots, \sigma_p$ are strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ and if $\gamma_{ik} \in \mathcal{C}_{\sigma_i} (1 \leq i \leq p, 1 \leq k \leq q_i)$ then, roughly speaking, a specialization of $(\gamma_{ik})_{1 \leq i \leq p,\, 1 \leq k \leq q_i}$ over $\mathcal{C}$ can, under certain general conditions, be extended to a specialization of $(\sigma_1, \cdots, \sigma_p)$ in such a way that various inequalities are preserved. An algebraico-geometric structure is introduced into $\mathfrak{G}^*$ in the following way. A subset $\mathfrak{M}^*$ of $\mathfrak{G}^*$ is called an *irreducible set* in $\mathfrak{G}^*$ if $\mathfrak{M}^*$ contains an element σ^* such that $\mathfrak{M}^*$ is the set of all specializations of σ^*; σ^* is then called a *generic element* of $\mathfrak{M}^*$, and the transcendence degree of $\mathcal{G}\langle \sigma^* \mathcal{G} \rangle$ over $\mathcal{G}$, which is the same as the transcendence degree of $\mathcal{C}_{\sigma^*}$ over $\mathcal{C}$ and does not depend on the choice of generic element σ^*, is called the *dimension* of $\mathfrak{M}^*$. A subset $\mathfrak{M}^*$ of $\mathfrak{G}^*$ is called an *algebraic set* in $\mathfrak{G}^*$ if $\mathfrak{M}^*$ is the union of a finite set of irreducible sets in $\mathfrak{G}^*$; the definition and elementary properties of the components of an algebraic set quickly follow. This algebraico-geometric structure in $\mathfrak{G}^*$ induces a similar structure in $\mathfrak{G}$. Some propositions are proved about algebraic sets in $\mathfrak{G}$ which are analogous to some elementary results in algebraic geometry. Finally, by combining the group structure and algebraico-geometric structure of $\mathfrak{G}$ we arrive at the concept of *algebraic group* in $\mathfrak{G}$. Several simple results about such algebraic groups are proved which are like certain known results on algebraic matric groups (Kolchin [3]) and, more generally, group varities in the sense of Weil [10].

In Chapter III, after a brief discussion of normal extensions of differential fields, the results of Chapter II are applied to develop a Galois theory of strongly normal extensions. It is shown that strong normality implies normality, but not conversely. Let $\mathcal{G}$ be a strongly normal extension of $\mathcal{F}$.

The group $\mathfrak{G}$ of all automorphisms of $\mathfrak{H}$ over $\mathfrak{F}$ is itself algebraic, and there is a one-to-one Galois correspondence between the set of all intermediate differential fields and a certain set of subgroups of $\mathfrak{G}$; this certain set is characterized as the set of all algebraic groups in $\mathfrak{G}$. The transcendence degree of $\mathfrak{H}$ over any intermediate differential field $\mathfrak{F}_1$ is proved to be equal to the dimension of the corresponding group $\mathfrak{G}(\mathfrak{F}_1)$; the component of $\mathfrak{G}$ containing the identity (which component is unique and is a normal algebraic subgroup of $\mathfrak{G}$ of finite index) corresponds to the relative algebraic closure of $\mathfrak{F}$ in $\mathfrak{H}$. It is shown that if $\mathfrak{F}_1$ is an intermediate differential field then the following conditions are equivalent: 1) $\mathfrak{F}_1$ is strongly normal over $\mathfrak{F}$; 2) $\mathfrak{F}_1$ is normal over $\mathfrak{F}$; 3) $\mathfrak{F}_1$ is weakly normal over $\mathfrak{F}$; 4) $\sigma\mathfrak{F}_1 \subseteq \mathfrak{F}_1$ for every $\sigma \in \mathfrak{G}$; 5) $\mathfrak{G}(\mathfrak{F}_1)$ is a normal subgroup of $\mathfrak{G}$. And when these conditions are satisfied, the factor group $\mathfrak{G}/\mathfrak{G}(\mathfrak{F}_1)$ is isomorphic with the group of all automorphisms of $\mathfrak{F}_1$ over $\mathfrak{F}$.

The remainder of Chapter III is devoted to three special types of extension. An element α is defined to be *primitive* over a differential field $\mathfrak{F}$ if $\delta_i\alpha \in \mathfrak{F}$ $(1 \leq i \leq m)$, to be *exponential* over $\mathfrak{F}$ if $\alpha \neq 0$ and $\alpha^{-1}\delta_i\alpha \in \mathfrak{F}$ $(1 \leq i \leq m)$, and to be *weierstrassian* over $\mathfrak{F}$ if α is not a constant, and there exist two elements $g_2, g_3 \in \mathfrak{C}$ with the polynomial $4y^3 - g_2y - g_3$ having simple roots only and m elements $a_1, \cdots, a_m \in \mathfrak{F}$ such that $(\delta_i\alpha)^2 = a_i^2(4\alpha^3 - g_2\alpha - g_3)$ $(1 \leq i \leq m)$. In all three cases, if the field of constants of $\mathfrak{F}\langle\alpha\rangle$ is $\mathfrak{C}$, $\mathfrak{F}\langle\alpha\rangle$ is strongly normal over $\mathfrak{F}$. In the first two cases $\mathfrak{F}\langle\alpha\rangle$ is a Picard-Vessiot extension of $\mathfrak{F}$, but in the third case it is not, unless it is algebraic; indeed, it can be shown that if α is weierstrassian over $\mathfrak{F}$ and if it is possible to find a family $\mathfrak{F}_0, \mathfrak{F}_1, \cdots, \mathfrak{F}_r$ of differential fields such that $\mathfrak{F}_0 = \mathfrak{F}$, $\mathfrak{F}_i$ is a Picard-Vessiot extension of $\mathfrak{F}_{i-1}$ $(1 \leq i \leq r)$, and $\alpha \in \mathfrak{F}_r$, then α is algebraic over $\mathfrak{F}$. Reciprocally, it is shown that if a Picard-Vessiot extension can be obtained by a sequence of adjunctions of algebraic, primitive, exponential, and weierstrassian elements, then it can be obtained by adjunction of algebraic, primitive, and exponential elements alone. When α is transcendental over $\mathfrak{F}$ then, in all three cases, $\mathfrak{F}\langle\alpha\rangle$ is of transcendence degree 1 over $\mathfrak{F}$. Conversely, it is proved that every strongly normal (and indeed every weakly normal) extension of $\mathfrak{F}$ of transcendence degree 1 can be obtained from $\mathfrak{F}$ by combining with algebraic adjunctions an adjunction of one of these three types. The proof of this converse, which is long and in places involves complicated computations, makes use of the well-known theorem that the group of automorphisms of an algebraic function field of one variable over an algebraically closed field of characteristic 0 is finite if the genus exceeds 1.

3. Problems. Various problems remain for investigation; we mention three, which are related.

First, there is the connection between algebraic groups of automorphisms as defined herein, and group varieties as defined by A. Weil. Is it always possible to identify the component of the identity of an algebraic group with a group variety? Conversely, is every group variety identifiable with an algebraic group?

Second, there is the task of characterizing, if possible, by algebraic-group properties, those strongly normal extensions which are Picard-Vessiot extensions.

Third, there is the significance of solvability, or even of commutativity, of the group of automorphisms of a strongly normal extension. For what sort of strongly normal extension is the group abelian? It is conceivable that investigation of this question will lead into the theory of abelian functions.

4. Notation. The notation used is more or less the same as in Kolchin [3], and is reasonably standard. We mention only that the degree of transcendence and the degree of differential transcendence of $\mathcal{G}$ over $\mathcal{F}$ are denoted by $\partial^{\circ}\mathcal{G}/\mathcal{F}$ and $\nabla^{\circ}\mathcal{G}/\mathcal{F}$ respectively.

Chapter I. Differential-algebraic preliminaries.

1. A lemma on polynomial ideals. Let K be a field of characteristic 0,[4] and let $y_1, \cdots, y_n$ be indeterminates. We shall prove the following lemma, which collects certain known facts in a form convenient for future use.

LEMMA. *Let $\mathfrak{p}$ be a prime ideal of $K[y_1, \cdots, y_n]$ of dimension d. For every extension L of K the ideal $L \cdot \mathfrak{p}$ generated by $\mathfrak{p}$ in $L[y_1, \cdots, y_n]$ is equal to its own radical; the minimal prime ideal divisors $\mathfrak{p}_1, \cdots, \mathfrak{p}_r$ of $L \cdot \mathfrak{p}$ all have dimension d; every generic zero of every $\mathfrak{p}_i$ is a generic zero of $\mathfrak{p}$, and every generic zero of $\mathfrak{p}$ is a zero of precisely one $\mathfrak{p}_i$. There exists, independent of L, an irreducible polynomial R with coefficients in K such that for every extension L of K the number of minimal prime ideal divisors of $L \cdot \mathfrak{p}$ equals the number of irreducible factors into which R splits over L.*

Proof. Let $\mathfrak{P}$ be the radical of $L \cdot \mathfrak{p}$, so that $\mathfrak{P} = \mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_r$ where $\mathfrak{p}_1, \cdots, \mathfrak{p}_r$ are the minimal prime ideal divisors of $L \cdot \mathfrak{p}$. Let $(\eta_{i1}, \cdots, \eta_{in})$ be a generic zero of $\mathfrak{p}_i$; $(\eta_{i1}, \cdots, \eta_{in})$ is obviously a zero of $\mathfrak{p}$. If

[4] This condition, which suffices for the present purposes, can be relaxed.

$F \in K[y_1, \cdots, y_n]$ vanishes at $(\eta_{i1}, \cdots, \eta_{in})$ then $F \in \mathfrak{p}_i$, so that if we let G be a polynomial in $L[y_1, \cdots, y_n]$ such that $G \in \bigcap_{j \neq i} \mathfrak{p}_j$, $G \notin \mathfrak{p}_i$ then $FG \in \mathfrak{P}$, whence for some exponent $e > 0$ we have $F^e G^e \in L \cdot \mathfrak{p}$. Therefore there exist elements $\lambda_k \in L$ linearly independent over K such that we may write $F^e G^e = \Sigma \lambda_k P_k$, where each $P_k \in \mathfrak{p}$, and $G^e = \Sigma \lambda_k G_k$, where each $G_k \in K[y_1, \cdots, y_n]$. From this we see that $\Sigma \lambda_k F^e G_k = \Sigma \lambda_k P_k$, so that $F^e G_k = P_k \in \mathfrak{p}$ for every k; since not every G_k belongs to $\mathfrak{p}$ (for otherwise G would belong to $\mathfrak{p}_i$) and since $\mathfrak{p}$ is prime we conclude that $F \in \mathfrak{p}$. This shows that $(\eta_{i1}, \cdots, \eta_{in})$ is a generic zero of $\mathfrak{p}$.

Let $(\eta_1, \cdots, \eta_n)$ be any generic zero of $\mathfrak{p}$. For the sake of definiteness we suppose that $\eta_1, \cdots, \eta_d$ are algebraically independent over K; then there exists an element ω such that $K(\eta_1, \cdots, \eta_n) = K(\eta_1, \cdots, \eta_d, \omega)$. Let w be a new indeterminate and let R be a polynomial in $K[y_1, \cdots, y_d, w]$ of as low degree as possible which vanishes at $(\eta_1, \cdots, \eta_d, \omega)$, so that R is irreducible over K. Because $(\eta_{i1}, \cdots, \eta_{in})$ is also a generic zero of $\mathfrak{p}$, $(\eta_{i1}, \cdots, \eta_{in})$ is a generic specialization [5] of $(\eta_1, \cdots, \eta_n)$ over K, and therefore can be extended to a generic specialization $(\eta_{i1}, \cdots, \eta_{in}, \omega_i)$ of $(\eta_1, \cdots, \eta_n, \omega)$ over K. Now $(\eta_{i1}, \cdots, \eta_{id}, \omega_i)$ is a zero of R and therefore of some irreducible factor of R over L; moreover $(\eta_{i1}, \cdots, \eta_{id}, \omega_i)$ is a zero of only one of these irreducible factors, for otherwise $(\eta_{i1}, \cdots, \eta_{id}, \omega_i)$ and therefore $(\eta_1, \cdots, \eta_d, \omega)$ would be a zero of $\partial R/\partial w$ which is of lower degree than R. We denote the irreducible factor of R over L which vanishes at $(\eta_{i1}, \cdots, \eta_{id}, \omega_i)$ by R_i.

Let $(\eta'_{i1}, \cdots, \eta'_{id}, \omega'_i)$ be a generic zero of the prime ideal $\mathfrak{r}_i$ generated by R_i in $L[y_1, \cdots, y_d, w]$. Then $(\eta_{i1}, \cdots, \eta_{id}, \omega_i)$ is a specialization of $(\eta'_{i1}, \cdots, \eta'_{id}, \omega'_i)$ over L and a generic specialization of $(\eta'_{i1}, \cdots, \eta'_{id}, \omega'_i)$ over K; therefore there exist elements $\eta'_{i,d+1}, \cdots, \eta'_{in}$ such that $(\eta_{i1}, \cdots, \eta_{in}, \omega_i)$ is a generic specialization of $(\eta'_{i1}, \cdots, \eta'_{in}, \omega'_i)$ over K. Now $(\eta'_{i1}, \cdots, \eta'_{in})$ is a zero of $\mathfrak{p}$ and therefore of $\mathfrak{p}_j$ for some j; since $(\eta_{i1}, \cdots, \eta_{in})$ must be a zero of this $\mathfrak{p}_j$ and since $\mathfrak{p}_j \not\subseteq \mathfrak{p}_i$ if $i \neq j$ it follows that $i = j$. As $\eta'_{i1}, \cdots, \eta'_{id}$ are obviously algebraically independent over L we have

$$\partial^0 L(\eta'_{i1}, \cdots, \eta'_{in})/L \geq d = \partial^0 K(\eta_{i1}, \cdots, \eta_{in})/K$$
$$\geq \partial^0 L(\eta_{i1}, \cdots, \eta_{in})/L = \dim \mathfrak{p}_i,$$

so that $(\eta'_{i1}, \cdots, \eta'_{in})$ is a generic zero of $\mathfrak{p}_i$ and $\dim \mathfrak{p}_i = d$. It follows that $(\eta'_{i1}, \cdots, \eta'_{in})$ is a generic specialization of $(\eta_{i1}, \cdots, \eta_{in})$ over L, so that

[5] The very elementary facts concerning specializations over a field used in this paper can be found in Weil [9], chapter II.

$(\eta'_{i1}, \cdots, \eta'_{in}, \omega'_i)$ is a generic specialization of $(\eta_{i1}, \cdots, \eta_{in}, \omega_i)$ over L, and $(\eta_{i1}, \cdots, \eta_{id}, \omega_i)$ is a generic zero of $\mathfrak{r}_i$.

If $R_i = R_j$ then $\mathfrak{r}_i = \mathfrak{r}_j$ and $(\eta_{i1}, \cdots, \eta_{id}, \omega_i)$ is a generic specialization of $(\eta_{j1}, \cdots, \eta_{jd}, \omega_j)$ over L, so that $(\eta_{i1}, \cdots, \eta_{in})$ is a generic specialization of $(\eta_{j1}, \cdots, \eta_{jn})$ over L, whence $\mathfrak{p}_i = \mathfrak{p}_j$ and $i = j$. Thus $R_1, \cdots, R_r$ are distinct irreducible factors of R over L. Let S be any irreducible factor of R over L, let $\mathfrak{s}$ denote the prime ideal generated by S in $L[y_1, \cdots, y_d, w]$, and let $(\zeta_1, \cdots, \zeta_d, \theta)$ be a generic zero of $\mathfrak{s}$; $(\zeta_1, \cdots, \zeta_d, \theta)$ is clearly a generic specialization of $(\eta_1, \cdots, \eta_d, \omega)$ over K and therefore can be extended to a generic specialization $(\zeta_1, \cdots, \zeta_n, \theta)$ of $(\eta_1, \cdots, \eta_n, \omega)$. $(\zeta_1, \cdots, \zeta_n)$ is a zero of $\mathfrak{p}$ and therefore of $\mathfrak{p}_i$ for some i; therefore $(\zeta_1, \cdots, \zeta_d, \theta)$ is a zero of R_i, so that R_i is divisible by S. It follows that $R_1, \cdots, R_r$ are all the irreducible factors of R over L, so that the number of minimal prime ideal divisors of $L \cdot \mathfrak{p}$ equals the number of irreducible factors of R over L. Also $(\eta_1, \cdots, \eta_n)$, which obviously is a zero of some $\mathfrak{p}_i$, is a zero of only one $\mathfrak{p}_i$; for if $(\eta_1, \cdots, \eta_n)$ were a zero of $\mathfrak{p}_i$ and $\mathfrak{p}_j$ $(i \neq j)$ then $(\eta_1, \cdots, \eta_d, \omega)$ would be a zero of R_i and R_j, and therefore of $\partial R/\partial w$, which is of lower degree than R.

It remains to prove that $\mathfrak{P} = L \cdot \mathfrak{p}$, and to do this it suffices to show that $\mathfrak{P} \subseteq L \cdot \mathfrak{p}$. If $F \in \mathfrak{P}$ then we may write $F = \Sigma \lambda_j F_j$, where each F_j belongs to $K[y_1, \cdots, y_n]$ and (λ_j) is a family of elements of L linearly independent over K. Let $A_{d+1}, B_{d+1}, \cdots, A_n, B_n$ be polynomials in $K[y_1, \cdots, y_d, w]$ such that

$$\eta_k = A_k(\eta_1, \cdots, \eta_d, \omega)/B_k(\eta_1, \cdots, \eta_d, \omega), \; d+1 \leq k \leq n.$$

Then there exists a single exponent $e \geq 0$ such that for each j

$$(B_{d+1} \cdots B_n)^e F_j \equiv G_j(B_{d+1} y_{d+1} - A_{d+1}, \cdots, B_n y_n - A_n),$$

where $G_j \in K[y_1, \cdots, y_d, w]$. It is easy to see that $\Sigma \lambda_j G_j$ vanishes at $(\eta_{i1}, \cdots, \eta_{id}, \omega_i)$ for each i and therefore is divisible by R. Because the λ_j's are linearly independent over K it easily follows that each G_j is divisible by R, so that each G_j vanishes at $(\eta_1, \cdots, \eta_d, \omega)$, each $(B_{d+1} \cdots B_n)^e F_j$ vanishes at $(\eta_1, \cdots, \eta_n, \omega)$, each F_j vanishes at $(\eta_1, \cdots, \eta_n)$, each $F_j \in \mathfrak{p}$, and $F \in L \cdot \mathfrak{p}$.

2. Prime differential ideals and differential field extension. Let $\mathfrak{F}$ be a differential field and let $y_1, \cdots, y_n$ denote indeterminates. If Π is a prime differential ideal of the differential ring $\mathfrak{F}\{y_1, \cdots, y_n\}$ and $(\eta_1, \cdots, \eta_n)$ is a generic zero of Π then the degree of differential transcendence

$\nabla^0\mathfrak{F}\langle\eta_1,\cdots,\eta_n\rangle/\mathfrak{F}$ is called the *dimension* of Π (notation: dim Π); dim Π does not depend on the particular generic zero used, and $0 \leq \dim \Pi \leq n$. The degree of transcendence $\partial^0\mathfrak{F}\langle\eta_1,\cdots,\eta_n\rangle/\mathfrak{F}$ is called the *order* of Π (notation: ord Π); ord Π, which does not depend on the particular generic zero used, either is an integer ≥ 0 (finite order) or is ∞ (infinite order). The following result is well-known for the case in which $\mathfrak{F}$ is an ordinary differential field (see e. g. Ritt [8], pp. 50-51).

PROPOSITION 1. *Let Π be a prime differential ideal of $\mathfrak{F}\{y_1,\cdots,y_n\}$. For every extension $\mathfrak{G}$ of $\mathfrak{F}$ the ideal $\mathfrak{G}\cdot\Pi$ generated by Π in $\mathfrak{G}\{y_1,\cdots,y_n\}$ is a perfect differential ideal; the minimal prime differential ideal divisors $\Pi_1,\cdots,\Pi_r$ of $\mathfrak{G}\cdot\Pi$ all have the same dimension as Π, and all have the same order as Π; every generic zero of every Π_i is a generic zero of Π, and every generic zero of Π is a zero of precisely one Π_i. There exists, independent of $\mathfrak{G}$, an irreducible polynomial R with coefficients in $\mathfrak{F}$ such that for every extension $\mathfrak{G}$ of $\mathfrak{F}$ the number of minimal prime differential ideal divisors of $\mathfrak{G}\cdot\Pi$ equals the number of irreducible factors into which R splits over $\mathfrak{G}$.*

Proof. Let $\mathfrak{R}$, $\mathfrak{R}'$ denote $\mathfrak{F}\{y_1,\cdots,y_n\}$, $\mathfrak{G}\{y_1,\cdots,y_n\}$ respectively. For each integer $k \geq 0$ let $\mathfrak{R}_k$, $\mathfrak{R}'_k$ denote the set of all elements of $\mathfrak{R}$, $\mathfrak{R}'$ respectively which do not have order $> k$. We shall consider $\mathfrak{R}_k$ and $\mathfrak{R}'_k$ as polynomial rings over $\mathfrak{F}$ and $\mathfrak{G}$ respectively (each element of $\mathfrak{R}_k$ and of $\mathfrak{R}'_k$ is a polynomial in the expressions $\delta_1^{i_1}\cdots\delta_m^{i_m}y_j$ with $0 \leq i_1+\cdots+i_m \leq k$, $1 \leq j \leq n$).

For every $k \geq 0$, $\Pi \cap \mathfrak{R}_k$ is a prime ideal of $\mathfrak{R}_k$, so that (§ 1) $\mathfrak{G}\cdot(\Pi\cap\mathfrak{R}_k)$ is an ideal of $\mathfrak{R}'_k$ which is equal to its own radical. Clearly $\mathfrak{G}\cdot\Pi$ is a differential ideal of $\mathfrak{R}'$; if $\mathfrak{G}\cdot\Pi$ were not perfect there would exist an $F \in \mathfrak{R}'$ and an integer $e > 0$ with $F \notin \mathfrak{G}\cdot\Pi$, $F^e \in \mathfrak{G}\cdot\Pi$, so that for large k we would have $F \notin \mathfrak{G}\cdot(\Pi\cap\mathfrak{R}_k)$, $F^e \in \mathfrak{G}\cdot(\Pi\cap\mathfrak{R}_k)$, contradicting the fact that $\mathfrak{G}\cdot(\Pi\cap\mathfrak{R}_k)$ is its own radical. Thus $\mathfrak{G}\cdot\Pi$ is a perfect differential ideal of $\mathfrak{R}'$.

We now assert that $\mathfrak{G}\cdot(\Pi\cap\mathfrak{R}_k) = (\mathfrak{G}\cdot\Pi)\cap\mathfrak{R}'_k$. Indeed, it is obvious that $\mathfrak{G}\cdot(\Pi\cap\mathfrak{R}_k) \subseteq (\mathfrak{G}\cdot\Pi)\cap\mathfrak{R}'_k$. Suppose then that $F \in (\mathfrak{G}\cdot\Pi)\cap\mathfrak{R}'_k$. Then we may write

$$F = \sum P_k\phi_k \tag{1}$$

where the elements $\phi_k \in \mathfrak{G}$ are linearly independent over $\mathfrak{F}$ and each $P_k \in \Pi$. Fixing our attention on any derivative $\delta_1^{i_1}\cdots\delta_m^{i_m}y_j$ of order $i_1+\cdots+i_m > k$, let C_k denote the coefficient in P_k of any fixed positive power of this derivative; since F is not of order $> k$, (1) yields the relation $\sum C_k\phi_k = 0$, so that each

$C_k = 0$. It follows that no P_k has order $> k$, so that each $P_k \in \mathfrak{R}_k$, and $F \in \mathfrak{G} \cdot (\Pi \cap \mathfrak{R}_k)$. This proves our assertion.

Since the perfect differential ideal $\mathfrak{G} \cdot \Pi$ is the intersection of its minimal prime differential ideal divisors $\Pi_1, \cdots, \Pi_r$ it is a consequence of the above assertion that

$$\mathfrak{G} \cdot (\Pi \cap \mathfrak{R}_k) = (\Pi_1 \cap \mathfrak{R}'_k) \cap \cdots \cap (\Pi_r \cap \mathfrak{R}'_k).$$

Now $\Pi_i \not\subseteq \Pi_j$ if $i \neq j$, so that if k is sufficiently great $\Pi_i \cap \mathfrak{R}'_k \not\subseteq \Pi_j \cap \mathfrak{R}'_k$ whenever $i \neq j$. Taking k large enough for this to be the case, we see that $\Pi_1 \cap \mathfrak{R}'_k, \cdots, \Pi_r \cap \mathfrak{R}'_k$ are the minimal prime ideal divisors of $\mathfrak{G} \cdot (\Pi \cap \mathfrak{R}_k)$ in $\mathfrak{R}'_k$.

It is obvious that Π_i is of dimension no higher than Π. If Π_i had lower dimension than Π then there would exist a subset $z_1, \cdots, z_t$ of $y_1, \cdots, y_n$ such that Π contains no nonzero differential polynomial in $z_1, \cdots, z_t$ alone but Π_i does, that is (for k large) $\Pi \cap \mathfrak{R}_k$ contains no nonzero polynomial in the expressions $\delta_1^{i_1} \cdots \delta_m^{i_m} z_l$ with $0 \leq i_1 + \cdots + i_m \leq k$, $1 \leq l \leq t$, but $\Pi_i \cap \mathfrak{R}'_k$ does; this would imply that the prime polynomial ideal $\Pi_i \cap \mathfrak{R}'_k$ of $\mathfrak{R}'_k$ has lower dimension than the prime polynomial ideal $\Pi \cap \mathfrak{R}_k$ of $\mathfrak{R}_k$, contradicting the lemma of § 1. Therefore each Π_i is of the same dimension as Π.

Again, the order of Π_i obviously equals the limit (including the possibility ∞) as k becomes infinite of the dimension of $\Pi_i \cap \mathfrak{R}'_k$, which by the lemma equals the limit of the dimension of $\Pi \cap \mathfrak{R}_k$, which equals the order of Π.

If $(\eta_{i1}, \cdots, \eta_{in})$ is a generic zero of Π_i then obviously

$$(\delta_1^{i_1} \cdots \delta_m^{i_m} \eta_{ij})_{0 \leq i_1 + \cdots + i_m \leq k,\ 1 \leq j \leq n}$$

is a generic zero of $\Pi_i \cap \mathfrak{R}'_k$ and therefore (by the lemma) a generic zero of $\Pi \cap \mathfrak{R}_k$; since k can be arbitrarily large this means that $(\eta_{i1}, \cdots, \eta_{in})$ is a generic zero of Π.

If $(\eta_1, \cdots, \eta_n)$ is a generic zero of Π then $(\eta_1, \cdots, \eta_n)$ is a zero of at least one Π_i; if it were a zero of Π_i for two distinct values of i then for k large $(\delta_1^{i_1} \cdots \delta_m^{i_m} \eta_j)_{0 \leq i_1 + \cdots + i_m \leq k,\ 1 \leq j \leq n}$ would be a generic zero of $\Pi \cap \mathfrak{R}_k$ and a zero of $\Pi_i \cap \mathfrak{R}'_k$ for two distinct values of i, contradicting the lemma.

It remains to prove the existence of a polynomial R as described in the statement of the proposition. To this end let $k(\mathfrak{G})$ be the smallest integer such that, for all $k \geq k(\mathfrak{G})$, $\Pi_1 \cap \mathfrak{R}'_k, \cdots, \Pi_r \cap \mathfrak{R}'_k$ are the minimal prime ideal divisors of $\mathfrak{G} \cdot (\Pi \cap \mathfrak{R}_k)$, that is such that, for all $k \geq k(\mathfrak{G})$, $\Pi_i \cap \mathfrak{R}'_k \not\subseteq \Pi_j \cap \mathfrak{R}'_k$ whenever $i \neq j$. We shall show below that $k(\mathfrak{G})$ is

an increasing function of $\mathfrak{G}$, that is, if $\mathfrak{H}$ is an extension of $\mathfrak{G}$ then $k(\mathfrak{G}) \leq k(\mathfrak{H})$. Assuming this result, let us see how we can complete the proof of the proposition. By the lemma for each $k \geq 0$ there exists, independent of $\mathfrak{G}$, an irreducible polynomial R_k with coefficients in $\mathfrak{F}$ such that the number of minimal prime ideal divisors of $\mathfrak{G} \cdot (\Pi \cap \mathfrak{R}_k)$ in $\mathfrak{R}'_k$ equals the number of irreducible factors of R_k over $\mathfrak{G}$. Now the number of irreducible factors of R_k over $\mathfrak{G}$ equals the number of irreducible factors of R_k over the relative algebraic closure $\mathfrak{F}^0$ of $\mathfrak{F}$ in $\mathfrak{G}$. Therefore

number of minimal prime differential ideal divisors of $\mathfrak{G} \cdot \Pi =$
number of minimal prime ideal divisors of $\mathfrak{G} \cdot (\Pi \cap \mathfrak{R}_k)$ (all $k \geq k(\mathfrak{G})$) =
number of irreducible factors of R_k over $\mathfrak{G}$ (all $k \geq k(\mathfrak{G})$) =
number of irreducible factors of R_k over $\mathfrak{F}^0$ (all $k \geq k(\mathfrak{G})$) =
number of irreducible factors of R_k over $\mathfrak{F}^0$ (all $k \geq k(\mathfrak{F}^0)$) =
number of irreducible factors of R_k over $\mathfrak{G}$ (all $k \geq k(\mathfrak{F}^0)$).

Thus if we let $\mathfrak{F}'$ be an algebraic closure [6] of $\mathfrak{F}$ and set $R = R_l$, where $l = k(\mathfrak{F}')$, then $l \geq k(\mathfrak{F}^0)$ and R does not depend on $\mathfrak{G}$, and the number of minimal prime differential ideal divisors of $\mathfrak{G} \cdot \Pi$ equals the number of irreducible factors of R over $\mathfrak{G}$.

We now show that $k(\mathfrak{G})$ is an increasing function of $\mathfrak{G}$. Let $\mathfrak{H}$ be an extension of $\mathfrak{G}$, let $\mathfrak{R}'' = \mathfrak{H}\{y_1, \cdots, y_n\}$, and let $\mathfrak{R}''_k$ denote the ring of all elements of $\mathfrak{R}''$ which do not have order $> k$. $\mathfrak{H} \cdot \Pi$ is a perfect differential ideal of $\mathfrak{R}''$ and $\mathfrak{H} \cdot \Pi = \mathfrak{H} \cdot (\mathfrak{G} \cdot \Pi) = \mathfrak{H} \cdot (\Pi_1 \cap \cdots \cap \Pi_r)$. Now if Π' is any ideal of R' and if for some $P'' \in \mathfrak{H} \cdot \Pi'$ we write $P'' = \sum \phi_l P'_l$, where each $P'_l \in \mathfrak{R}'$ and the elements ϕ_l of $\mathfrak{H}$ are linearly independent over $\mathfrak{G}$, then it is easy to see that each $P'_l \in \Pi'$. It follows that

$$(\mathfrak{H} \cdot \Pi_1) \cap \cdots \cap (\mathfrak{H} \cdot \Pi_r) = \mathfrak{H} \cdot (\Pi_1 \cap \cdots \cap \Pi_r),$$

so that $\mathfrak{H} \cdot \Pi = (\mathfrak{H} \cdot \Pi_1) \cap \cdots \cap (\mathfrak{H} \cdot \Pi_r)$. Now if $i_1 \neq i_2$ then a minimal prime differential ideal divisor Λ_1 of $\mathfrak{H} \cdot \Pi_{i_1}$ can not be contained in a minimal prime differential ideal divisor Λ_2 of $\mathfrak{H} \cdot \Pi_{i_2}$, because otherwise for every k we would have $\Lambda_2 \cap \mathfrak{R}''_k \supseteq \mathfrak{H} \cdot (\Pi_{i_1} \cap \mathfrak{R}'_k + \Pi_{i_2} \cap \mathfrak{R}'_k)$ and $\Lambda_2 \cap \mathfrak{R}''_k$

[6] If K is any field of characteristic 0 and K' is an algebraic closure of K then every derivation of K has a unique extension which is a derivation of K' (see, for example, Bourbaki [1], chapter V, § 9, proposition 5, p. 139); moreover, it is easy to verify that if two derivations of K commute then there extended derivations of K' commute. It follows that every differential field $\mathfrak{f}$ has an algebraically closed algebraic differential field extension; we call any such extension an *algebraic closure* of the differential field $\mathfrak{f}$. Any two algebraic closures of $\mathfrak{f}$ are isomorphic over $\mathfrak{f}$.

would have lower dimension than $\Pi_{i_2} \cap \mathfrak{R}'_k$ in contradiction to the lemma of § 1 (since for k large $\Lambda_2 \cap \mathfrak{R}''_k$ is a minimal prime ideal divisor of $\mathfrak{A} \cdot (\Pi_{i_2} \cap \mathfrak{R}'_k)$). Therefore if we denote the minimal prime differential ideal divisors of $\mathfrak{A} \cdot \Pi_i$ by $\Pi_{i1}, \cdots, \Pi_{i,s(i)}$ $(1 \leq i \leq r)$ then the minimal prime differential ideal divisors of $\mathfrak{A} \cdot \Pi$ are the ideals Π_{ij} $(1 \leq i \leq r,$ $1 \leq j \leq s(i))$. If $k < k(\mathfrak{G})$ then there exist i, i' with $i \neq i'$ such that $\Pi_i \cap \mathfrak{R}'_k \subseteq \Pi_{i'} \cap \mathfrak{R}'_k$; for these i, i' we have

$$(\Pi_{i1} \cap \mathfrak{R}''_k) \cap \cdots \cap (\Pi_{i,s(i)} \cap \mathfrak{R}''_k) = (\mathfrak{A} \cdot \Pi_i) \cap \mathfrak{R}''_k$$
$$\subseteq (\mathfrak{A} \cdot \Pi_{i'}) \cap \mathfrak{R}''_k \subseteq \Pi_{i'1} \cap \mathfrak{R}''_k,$$

so that for some j we have $\Pi_{ij} \cap \mathfrak{R}''_k \subseteq \Pi_{i'1} \cap \mathfrak{R}''_k$, whence $k < k(\mathfrak{A})$. It follows that $k(\mathfrak{G}) \leq k(\mathfrak{A})$. As we have seen, this completes the proof of Proposition 1.

3. Specializations over differental fields. For purposes of convenience we extend the language of specializations, as used in algebraic geometry, to differential fields. Let $\mathfrak{F}$ be a differential field and let $(\eta_j)_{j \in J}$ be an indexed family of elements of some extension of $\mathfrak{F}$. A family $(\zeta_j)_{j \in J}$, with the same set of indices J, of elements of some extension of $\mathfrak{F}$ will be called a *specialization of* $(\eta_j)_{j \in J}$ *over* $\mathfrak{F}$ if, for every finite subset $j_1, \cdots, j_n$ of J, every differential polynomial in $\mathfrak{F}\{y_1, \cdots, y_n\}$ which vanishes at $(\eta_{j_1}, \cdots, \eta_{j_n})$ also vanishes at $(\zeta_{j_1}, \cdots, \zeta_{j_n})$. If $(\zeta_j)_{j \in J}$ is a specialization of $(\eta_j)_{j \in J}$ over $\mathfrak{F}$ such that $(\eta_j)_{j \in J}$ is a specialization of $(\zeta_j)_{j \in J}$ over $\mathfrak{F}$ then we say that $(\zeta_j)_{j \in J}$ is a *generic* specialization of $(\eta_j)_{j \in J}$ over $\mathfrak{F}$. If I is a subset of J and $(\zeta_j)_{j \in J}$ is a specialization of $(\eta_j)_{j \in J}$ over $\mathfrak{F}$ then $(\zeta_i)_{i \in I}$ is a specialization of $(\eta_i)_{i \in I}$ over $\mathfrak{F}$; we say in this case that the specialization $(\zeta_j)_{j \in J}$ of $(\eta_j)_{j \in J}$ over $\mathfrak{F}$ is an *extension* of the specialization $(\zeta_i)_{i \in I}$ of $(\eta_i)_{i \in I}$ over $\mathfrak{F}$. If $(\zeta_j)_{j \in J}$ is a generic specialization of $(\eta_j)_{j \in J}$ over $\mathfrak{F}$ then there exists a unique isomorphism of $\mathfrak{F}\langle(\eta_j)_{j \in J}\rangle$ onto $\mathfrak{F}\langle(\zeta_j)_{j \in J}\rangle$ over $\mathfrak{F}$ which maps η_j onto ζ_j for every $j \in J$. If $(\zeta_j)_{j \in J}$ is a generic specialization of $(\eta_j)_{j \in J}$ over $\mathfrak{F}$ and if $(\eta'_{j'})_{j' \in J'}$ is any family of elements of some extension of $\mathfrak{F}\langle(\eta_j)_{j \in J}\rangle$, then the specialization can be extended to a generic specialization

$$((\zeta_j)_{j \in J}, (\zeta'_{j'})_{j' \in J'}) \text{ of } ((\eta_j)_{j \in J}, (\eta'_{j'})_{j' \in J'})$$

over $\mathfrak{F}$. The following proposition is well-known in the case of ordinary differential fields (Ritt [8], p. 49).

PROPOSITION 2. *If* $(\zeta_j)_{j \in J}$ *is a specialization of* $(\eta_j)_{j \in J}$ *over* $\mathfrak{F}$ *then*

$$\nabla^0 \mathfrak{F}\langle(\zeta_j)_{j \in J}\rangle/\mathfrak{F} \leq \nabla^0 \mathfrak{F}\langle(\eta_j)_{j \in J}\rangle/\mathfrak{F}$$

and

$$\partial^0\mathfrak{F}\langle(\zeta_j)_{j\in J}\rangle/\mathfrak{F} \leq \partial^0\mathfrak{F}\langle(\eta_j)_{j\in J}\rangle/\mathfrak{F};$$

if in addition $\partial^0\mathfrak{F}\langle(\eta_j)_{j\in J}\rangle/\mathfrak{F}$ *is finite and equal to* $\partial^0\mathfrak{F}\langle(\zeta_j)_{j\in J}\rangle/\mathfrak{F}$ *then the specialization is generic.*

Proof. The first part of the proposition is obvious. We prove the second part. Since $\partial^0\mathfrak{F}\langle(\eta_j)_{j\in J}\rangle/\mathfrak{F}$ and $\partial^0\mathfrak{F}\langle(\zeta_j)_{j\in J}\rangle/\mathfrak{F}$ are finite and equal there exist a finite subset K of J such that

$$\partial^0\mathfrak{F}\langle(\zeta_j)_{j\in J'}\rangle/\mathfrak{F} = \partial^0\mathfrak{F}\langle(\zeta_j)_{j\in J}\rangle/\mathfrak{F} = \partial^0\mathfrak{F}\langle(\eta_j)_{j\in J}\rangle/\mathfrak{F} = \partial^0\mathfrak{F}\langle(\eta_j)_{j\in J'}\rangle/\mathfrak{F}$$

for every finite subset J' of J which contains K. It is clear that $(\zeta_j)_{j\in J}$ is a generic specialization of $(\eta_j)_{j\in J}$ if $(\zeta_j)_{j\in J'}$ is a generic specialization of $(\eta_j)_{j\in J'}$ for every finite subset J' of J which contains K. It follows that we may assume that J is finite. Making this assumption, it is easy to see that there exists an integer $k_0 \geq 0$ such that

$$\mathfrak{F}\langle(\eta_j)_{j\in J}\rangle = \mathfrak{F}\left((\delta_1^{i_1}\cdots\delta_m^{i_m}\eta_j)_{0\leq i_1+\ldots+i_m\leq k,\, j\in J}\right),$$
$$\mathfrak{F}\langle(\zeta_j)_{j\in J}\rangle = \mathfrak{F}\left((\delta_1^{i_1}\cdots\delta_m^{i_m}\zeta_j)_{0\leq i_1+\ldots+i_m\leq k,\, j\in J}\right)$$

for every integer $k \geq k_0$. By a well-known result concerning specializations over a field (for example, see Weil [9], p. 28, Theorem 3) it follows that if we regard $\mathfrak{F}$ as a field then $(\delta_1^{i_1}\cdots\delta_m^{i_m}\zeta_j)_{0\leq i_1+\ldots+i_m\leq k,\, j\in J}$ is a generic specialization of $(\delta_1^{i_1}\cdots\delta_m^{i_m}\eta_j)_{0\leq i_1+\ldots+i_m\leq k,\, j\in J}$ over $\mathfrak{F}$ for every $k \geq k_0$. This implies that $(\zeta_j)_{j\in J}$ is a generic specialization of $(\eta_j)_{j\in J}$ over the differential field $\mathfrak{F}$.

COROLLARY. *A zero* $(\zeta_1, \cdots, \zeta_n)$ *of a prime differential ideal* Π *of* $F\{y_1, \cdots, y_n\}$ *of finite order is generic if and only if*

$$\partial^0\mathfrak{F}\langle\zeta_1, \cdots, \zeta_n\rangle/\mathfrak{F} = \operatorname{ord}\Pi.$$

4. Constants. Let $\mathfrak{F}$ be a differential field, and denote the field of constants of $\mathfrak{F}$ by $\mathfrak{C}$. By the *order* of a differential operator $\delta_1^{i_1}\cdots\delta_m^{i_m}$ we mean the inteer $i_1 + \cdots + i_m$. If $\theta_1, \cdots, \theta_n$ are differential operators of the form $\delta_1^{i_1}\cdots\delta_m^{i_m}$ $(0 \leq i_1 < \infty, \cdots, 0 \leq i_m < \infty)$ then we use $W_{\theta_1\ldots\theta_n}$ to denote the differential polynomial defined by $W_{\theta_1\ldots\theta_n} = \det(\theta_i y_j)$.

PROPOSITION 3. *The elements* $\eta_1, \cdots, \eta_n$ *of* $\mathfrak{F}$ *are linearly dependent over* $\mathfrak{C}$ *if and only if* $W_{\theta_1\ldots\theta_n}(\eta_1, \cdots, \eta_n) = 0$ *for all choices of* $\theta_1, \cdots, \theta_n$ *of order* $< n$.

This has been proved in Kolchin [6].

A consequence of this proposition is that if $\eta_1, \cdots, \eta_n$ are linearly dependent (or independent) over the field of constants of some differential field containing them then they are linearly dependent (or independent) over the field of constants of any differential field containing them; therefore we may speak simply of linear dependence or independence *over constants.*

COROLLARY 1. *Let L be a homogeneous linear polynomial in $\mathfrak{F}[u_1, \cdots, u_q]$. There exist a finite number of homogeneous linear polynomials $L_1, \cdots, L_r$ in $\mathfrak{k}[u_1, \cdots, u_q]$ such that q constants in an extension of $\mathfrak{F}$ form a zero of L if and only if they form a zero of $L_1, \cdots, L_r$.*

Proof. Write $L = \sum_{i=1}^{r} L_i \alpha_i$, where $\alpha_1, \cdots, \alpha_r$ are elements of $\mathfrak{F}$ linearly independent over $\mathfrak{k}$ (and therefore over constants) and each L_i is a homogeneous linear polynomial in $\mathfrak{k}[u_1, \cdots, u_q]$. If $\gamma_1, \cdots, \gamma_q$ are constants then so is $L_i(\gamma_1, \cdots, \gamma_q)$, $1 \leq i \leq r$, so that $L(\gamma_1, \cdots, \gamma_q) = 0$ if and only if $L_i(\gamma_1, \cdots, \gamma_q) = 0$, $1 \leq i \leq r$.

COROLLARY 2. *Let $\mathfrak{m}'$ be a set of polynomials in $\mathfrak{F}[u_1, \cdots, u_q]$. There exists a set $\mathfrak{m}$ of polynomials in $\mathfrak{k}[u_1, \cdots, u_q]$ such that q constants in an extension of $\mathfrak{F}$ form a zero of $\mathfrak{m}'$ if and only if they form a zero of $\mathfrak{m}$.*

Proof. This follows from Corollary 1 since each polynomial in $u_1, \cdots, u_q$ is a linear combination of power products in $u_1, \cdots, u_q$.

COROLLARY 3. *Let $\gamma_1, \cdots, \gamma_q$ be constants in an extension of $\mathfrak{F}$. Then $\partial^0 \mathfrak{F}\langle \gamma_1, \cdots, \gamma_q \rangle / \mathfrak{F} = \partial^0 \mathfrak{k}\langle \gamma_1, \cdots, \gamma_q \rangle / \mathfrak{k}$.*

Proof. By Corollary 2, $\gamma_{i_1}, \cdots, \gamma_{i_d}$ are algebraically dependent over $\mathfrak{F}$ if and only if they are over $\mathfrak{k}$.

COROLLARY 4. *If $\mathfrak{G}$ is an extension of $\mathfrak{F}$ with field of constants $\mathfrak{D}$ and $\mathfrak{E}$ is a differential subfield of $\mathfrak{F}$ then $\mathfrak{E}\langle \mathfrak{D} \rangle \cap \mathfrak{F} = \mathfrak{E}\langle \mathfrak{k} \rangle$.*

Proof. If $\alpha \in \mathfrak{E}\langle \mathfrak{k} \rangle$ then obviously $\alpha \in \mathfrak{E}\langle \mathfrak{D} \rangle \cap \mathfrak{F}$. Conversely, let $\alpha \in \mathfrak{E}\langle \mathfrak{D} \rangle \cap \mathfrak{F}$. Then there exist elements $e_1, \cdots, e_r \in \mathfrak{E}$ linearly independent over constants and elements $d_1, \cdots, d_r, d'_1, \cdots, d'_r \in \mathfrak{D}$ with $d'_1, \cdots, d'_r$ not all 0 (so that $\sum d'_i e_i \neq 0$) such that $\alpha = \sum d_i e_i / \sum d'_i e_i$, that is $\sum d'_i e_i \alpha - \sum d_i e_i = 0$. This means that $e_1\alpha, \cdots, e_r\alpha, e_1, \cdots, e_r$ are linearly dependent over constants, and therefore over $\mathfrak{k}$; thus there exist elements $c_1, \cdots, c_r, c'_1, \cdots, c'_r \in \mathfrak{k}$ not all 0 such that $\sum c'_i e_i \alpha - \sum c_i e_i = 0$. Since $e_1, \cdots, e_r$ are linearly independent over constants it follows that $c'_1, \cdots, c'_r$ are not all 0, so that $\sum c'_i e_i \neq 0$, whence $\alpha = \sum c_i e_i / \sum c'_i e_i \in \mathfrak{E}\langle \mathfrak{k} \rangle$.

COROLLARY 5. *Let $\mathcal{D}$ be a field of constants containing $\mathcal{C}$ and contained in some extension of $\mathcal{F}$. Then the field of constants of $\mathcal{F}\langle\mathcal{D}\rangle$ is $\mathcal{D}$.*

Proof. Let e be a nonzero constant in $\mathcal{F}\langle\mathcal{D}\rangle$. Then we may write $e\sum_{j=1}^{s}\alpha''_j d''_j = \sum_{i=1}^{r}\alpha'_i d'_i$, where $d'_i, d''_j \in \mathcal{D}$, $\alpha'_i, \alpha''_j \in \mathcal{F}$, $\sum_{j=1}^{s}\alpha''_j d''_j \neq 0$. We suppose that of all such equations ours is one for which s is minimal, and also, without loss of generality, that $\alpha''_s = 1$. For each $k (1 \leq k \leq m)$, $e\sum_{j=1}^{s-1}(\delta_k\alpha''_j)d''_j = \sum_{i=1}^{r}(\delta_k\alpha'_i)d'_i$, and this would contradict the minimal nature of s unless $\sum_{j=1}^{s}(\delta_k\alpha''_j)d''_j = 0$; thus $\sum_{j=1}^{s}\alpha''_j d''_j$ is a constant. It follows that our corollary will be proved if we show that every nonzero constant $a \in \mathcal{F}\langle\mathcal{D}\rangle$ of the form $a = \sum_{i=1}^{q}\alpha_i d_i$ $(\alpha_i \in \mathcal{F}, d_i \in \mathcal{D})$ belongs to $\mathcal{D}$. To this end we may suppose that $\alpha_1, \cdots, \alpha_q$ are linearly independent over constants and that every $d_i \neq 0$. Since $\sum_{i=1}^{q}(\delta_k\alpha_i)d_i = \delta_k a = 0$ $(1 \leq k \leq m)$ it follows from Corollary 1 that there exist elements $c_1, \cdots, c_q$ of $\mathcal{C}$ not all 0 such that $\sum_{i=1}^{q}(\delta_k\alpha_i)c_i = 0$ $(1 \leq k \leq m)$, so that the element $c = \sum_{i=1}^{q}\alpha_i c_i$ belongs to $\mathcal{C}$. Now, $\alpha_1, \cdots, \alpha_q$ are linearly independent and

$$\sum_{i=1}^{q}\alpha_i(ac_i - cd_i) = a\sum_{i=1}^{q}\alpha_i c_i - c\sum_{i=1}^{q}\alpha_i d_i = 0,$$

so that each $ac_i - cd_i = 0$, whence $a \in \mathcal{D}$.

5. Universal extensions. Let $\mathcal{F}^*$ be a differential field and let $\mathcal{F}$ be a differential subfield of $\mathcal{F}^*$. We shall call $\mathcal{F}^*$ a *universal extension* of $\mathcal{F}$ if, for every finitely generated differential field extension $\mathcal{F}_1$ of $\mathcal{F}$ with $\mathcal{F}_1 \subseteq \mathcal{F}^*$ and every integer $n > 0$ and every prime differential ideal Π of $\mathcal{F}_1\{y_1, \cdots, y_n\}$ not containing 1, there exists a generic zero $(\eta_1, \cdots, \eta_n)$ of Π with $\eta_1, \cdots, \eta_n \in \mathcal{F}^*$. A necessary and sufficient condition for an extension $\mathcal{F}^*$ of $\mathcal{F}$ to be universal is that for every finitely generated extension $\mathcal{F}_1$ of $\mathcal{F}$ with $\mathcal{F}_1 \subseteq \mathcal{F}^*$ and every finitely generated extension $\mathcal{G}$ of $\mathcal{F}_1$ there exist an isomorphim of $\mathcal{G}$ over $\mathcal{F}_1$ into $\mathcal{F}^*$ (that is, an isomorphism σ of $\mathcal{G}$ into $\mathcal{F}^*$ such that $\sigma a = a$ for every $a \in \mathcal{F}_1$). If $\mathcal{F}^*$ is a universal extension of $\mathcal{F}$ then $\mathcal{F}^*$ is a universal extension of every finitely generated extension of $\mathcal{F}$ contained in $\mathcal{F}^*$, and $\mathcal{F}^*$ is also a universal extension of every differential subfield of $\mathcal{F}$. If $\mathcal{F}^*$ is a universal extension of $\mathcal{F}$ then the degree of differential transcendence of $\mathcal{F}^*$ over $\mathcal{F}$ is infinite,

and (because of the Ritt basis theorem) for every integer $n > 0$ every differential ideal in $\mathfrak{F}^*\{y_1, \cdots, y_n\}$ not containing 1 has a zero $(\eta_1, \cdots, \eta_n)$ with $\eta_1, \cdots, \eta_n \in \mathfrak{F}^*$; in particular $\mathfrak{F}^*$ is algebraically closed, and the field of constants of $\mathfrak{F}^*$ is an alegbraically closed extension of the field of constants of $\mathfrak{F}$ of infinite degree of transcendence.

We shall prove below that every differential field $\mathfrak{F}$ has a universal extension $\mathfrak{F}^*$. Once this fact is known it is possible to define the *manifold* of a set Φ of differential polynomials in $\mathfrak{F}\{y_1, \cdots, y_n\}$ as the set of all zeros $(\eta_1, \cdots, \eta_n)$ of Φ with $\eta_1, \cdots, \eta_n \in \mathfrak{F}^*$; this use of universal extensions extends the well-known procedure of modern algebraic geometry, and gives a workable definition of manifold free of the logical difficulty involved in using "the set of all extension of $\mathfrak{F}$" (see Ritt [8], footnote 2 on p. 21).

Let Π be a prime differential ideal of $\mathfrak{F}\{y_1, \cdots, y_n\}$. If $\mathfrak{G}$ is an extension of $\mathfrak{F}$ the ideal $\mathfrak{G} \cdot \Pi$ of $\mathfrak{G}\{y_1, \cdots, y_n\}$ is a perfect differential ideal (Proposition 1); we shall say that Π is *absolutely prime* if $\mathfrak{G} \cdot \Pi$ is prime for every extension $\mathfrak{G}$. If $\mathfrak{G} \cdot \Pi$ is prime when we take for $\mathfrak{G}$ some algebraically closed extension of $\mathfrak{F}$ then Π is absolutely prime (because of Proposition 1 and the fact that every polynomial R over $\mathfrak{F}$ which is irreducible over an algebraically closed extension of $\mathfrak{F}$ is absolutely irreducible). In particular, if $\mathfrak{F}$ is algebraically closed then every prime differential ideal of $\mathfrak{F}\{y_1, \cdots, y_n\}$ is absolutely prime.

PROPOSITION 4. *Let I be a nonempty set of indices; for each $i \in I$ let n_i be an integer > 0; let $(y_{ij})_{i \in I, 1 \leq j \leq n_i}$ be a family of indeterminates; for each $i \in I$ let Π_i be an absolutely prime differential ideal of $\mathfrak{F}\{y_{i1}, \cdots, y_{in_i}\}$ not containing* 1. *Then the ideal Π generated by $\bigcup_{i \in I} \Pi_i$ in $\mathfrak{F}\{(y_{ij})_{i \in I, 1 \leq j \leq n_i}\}$ is a prime differential ideal not containing* 1. *If I is finite then Π is absolutely prime and* $\operatorname{ord} \Pi = \sum_{i \in I} \operatorname{ord} \Pi_i$.

Proof. That Π is a differential ideal is obvious. To prove that Π is prime and $1 \notin \Pi$ it suffices to consider the case in which I is finite; by induction then the entire proposition can be reduced to the case in which I consists of two elements. Accordingly, let I consist of the numbers 1 and 2. Then Π consists of all differential polynomials P which can be writen in the form

$$(2) \qquad P = \sum_{k_1} C_{2k_1} P_{1k_1} + \sum_{k_2} C_{1k_2} P_{2k_2} (P_{ik_i} \in \Pi_i, C_{ik} \in \mathfrak{F}\{y_{i1}, \cdots, y_{in}\}).$$

Let $(\eta_{11}, \cdots, \eta_{1n_1})$ be a generic zero of Π_1. Since Π_2 is absolutely prime the ideal Λ_2 generated by Π_2 in $\mathfrak{F}\langle \eta_{11}, \cdots, \eta_{1n_1} \rangle \{y_{21}, \cdots, y_{2n_2}\}$ is a prime differential ideal, and obviously $1 \notin \Lambda_2$; let $(\eta_{21}, \cdots, \eta_{2n_2})$ be a generic zero

of Λ_2. We shall show that $(\eta_{11}, \cdots, \eta_{1n_1}, \eta_{21}, \cdots, \eta_{2n_2})$ is a generic zero of Π, thereby proving that Π is prime, that $1 \notin \Pi$, and that

$$\begin{aligned}\operatorname{ord} \Pi &= \partial^0 \mathfrak{F}\langle \eta_{11}, \cdots, \eta_{1n}, \eta_{21}, \cdots, \eta_{2n_2}\rangle/\mathfrak{F} \\ &= \partial^0 \mathfrak{F}\langle \eta_{11}, \cdots, \eta_{1n_1}, \eta_{21}, \cdots, \eta_{2n_2}\rangle/\mathfrak{F}\langle \eta_{11}, \cdots, \eta_{1n_1}\rangle \\ &\quad + \partial^0 \mathfrak{F}\langle \eta_{11}, \cdots, \eta_{1n_1}\rangle/\mathfrak{F} = \operatorname{ord} \Lambda_2 + \operatorname{ord} \Pi_1 = \operatorname{ord} \Pi_2 + \operatorname{ord} \Pi_1.\end{aligned}$$

It is clear from (2) that $(\eta_{11}, \cdots, \eta_{1n_1}, \eta_{21}, \cdots, \eta_{2n_2})$ is a zero of Π. Let P be any differential polynomial in $\mathfrak{F}\{y_{11}, \cdots, y_{1n_1}, y_{21}, \cdots, y_{2n_2}\}$ which vanishes at $(\eta_{11}, \cdots, \eta_{1n_1}, \eta_{21}, \cdots, \eta_{2n_2})$. Then

$$P(\eta_{11}, \cdots, \eta_{1n_1}, y_{21}, \cdots, y_{2n_2}) \in \Lambda_2.$$

We now write

$$\begin{aligned}&P(\eta_{11}, \cdots, \eta_{1n_1}, y_{21}, \cdots, y_{2n_2}) \\ &\qquad = \sum_{k_2} C_{1k_2}(\eta_{11}, \cdots, \eta_{1n_1}) P_{2k_2}(y_{21}, \cdots, y_{2n_2}),\end{aligned}$$

where

$$C_{1k_2}(y_{11}, \cdots, y_{1n_1}) \in \mathfrak{F}\{y_{11}, \cdots, y_{1n_1}\},\ P_{2k_2}(y_{21}, \cdots, y_{2n_2}) \in \mathfrak{F}\{y_{21}, \cdots, y_{2n_2}\},$$

and the elements $C_{1k_2}(\eta_{11}, \cdots, \eta_{1n_1})$ of $\mathfrak{F}\langle \eta_{11}, \cdots, \eta_{1n_1}\rangle$ are linearly independent over $\mathfrak{F}$; it is easy to see, since $\Lambda_2 = \mathfrak{F}\langle \eta_{11}, \cdots, \eta_{1n_1}\rangle \cdot \Pi_2$, that each $P_{2k_2} \in \Pi_2$. Let

$$Q = P - \sum_{k_2} C_{1k_2} P_{2k_2},$$

so that $Q(\eta_{11}, \cdots, \eta_{1n_1}, y_{21}, \cdots, y_{2n_2}) = 0$; if we write

$$Q = \sum_{k_1} C_{2k_1} P_{1k_1},$$

where the C_{2k_1} are distinct power products in $y_{21}, \cdots, y_{2n_2}$ and their derivatives of various orders, and each $P_{1k_1} \in \mathfrak{F}\{y_{11}, \cdots, y_{1n_1}\}$, then each $P_{1k_1}(\eta_{11}, \cdots, \eta_{1n_1}) = 0$, so that each $P_{1k_1} \in \Pi_1$. It follows that P can be written in the form (2) and therefore belongs to Π; therefore $(\eta_{11}, \cdots, \eta_{1n_1}, \eta_{21}, \cdots, \eta_{2n_2})$ is a generic zero of Π.

To complete the proof it remains to show that Π is absolutely prime. To this end let $\mathfrak{G}$ be any extension of $\mathfrak{F}$. Clearly $\mathfrak{G} \cdot \Pi$ is the ideal generated by $(\mathfrak{G} \cdot \Pi_1) \cup (\mathfrak{G} \cdot \Pi_2)$ in $\mathfrak{G}\{y_{11}, \cdots, y_{1n_1}, y_{21}, \cdots, y_{2n_2}\}$, and $\mathfrak{G} \cdot \Pi$, $\mathfrak{G} \cdot \Pi_2$ are prime (because Π_1, Π_2 are absolutely prime); therefore by what we have already proved $\mathfrak{G} \cdot \Pi$ is prime. Thus Π is absolutely prime.

REMARK. We observe from the proof that the hypothesis in Proposition 4 that each Π_i be absolutely prime may be weakened. It is enough to assume that, for each $i \in I$, Π_i is prime and $\mathfrak{F}_i \cdot \Pi_i$ is prime whenever $\mathfrak{F}_i$ is an

extension of $\mathfrak{F}$ obtained by the adjunction of generic zeros of a finite number of Π_j's with $j \neq i$. Except for the statement that Π is *absolutely* prime, the conclusion of Proposition 4 is then valid (Π still being prime).

THEOREM. *Every differential field has a universal extension.*

Proof. We lose no generality in assuming that the given differential field $\mathfrak{F}$ is algebraically closed, for a universal extension of an algebraic closure of $\mathfrak{F}$ is a universal extension of $\mathfrak{F}$. We shall show that for every algebraically closed differential field $\mathfrak{G}$ there exists an extension $\mathfrak{G}^\dagger$ of $\mathfrak{G}$ with the following two properties: 1) $\mathfrak{G}^\dagger$ is algebraically closed; 2) for every integer $n > 0$ and for every prime differential ideal Π of $\mathfrak{G}\{y_1, \cdots, y_n\}$ not containing 1 there exists a generic zero $(\eta_1, \cdots, \eta_n)$ of Π with $\eta_1, \cdots, \eta_n \in \mathfrak{G}^\dagger$. Once this is done we can define inductively a sequence of differential fields $\mathfrak{F}^{(k)}$ such that $\mathfrak{F}^{(0)} = \mathfrak{F}$ and $\mathfrak{F}^{(k+1)} = \mathfrak{F}^{(k)\dagger}$ for every integer $k \geq 0$; the union $\mathfrak{F}^* = \cup \mathfrak{F}^{(k)}$ will then be a differential field which, as is easy to see, is a universal extension of $\mathfrak{F}$.

Let $\mathfrak{P}_n$ be the set of all prime differential ideals in $\mathfrak{G}\{y_1, \cdots, y_n\}$ which do not contain 1; since $\mathfrak{G}$ is algebraically closed, every element of $\mathfrak{P}_n$ is absolutely prime. Let $(y_{n\Pi j})_{1 \leq n < \infty, \Pi \in \mathfrak{P}_n, 1 \leq j \leq n}$ be a family of indeterminates. For each $\Pi \in \mathfrak{P}_n$ let $\Lambda(n, \Pi)$ denote the set which is obtained when in all the differential polynomials in Π we replace y_j by $y_{n\Pi j}$ $(1 \leq j \leq n)$; $\Lambda(n, \Pi)$ is obviously an absolutely prime differential ideal of $\mathfrak{G}\{y_{n\Pi 1}, \cdots, y_{n\Pi n}\}$ which does not contain 1. It follows from Proposition 4 that the ideal Λ generated by $\bigcup_{1 \leq n < \infty, \Pi \in \mathfrak{P}_n} \Lambda(n, \Pi)$ in the differential ring $\mathfrak{R} = \mathfrak{G}\{(y_{n\Pi j})_{1 \leq n < \infty, \Pi \in \mathfrak{P}_n, 1 \leq j \leq n}\}$ is a prime differential ideal not containing 1. The differential ring of residue classes $\mathfrak{R}/\Lambda$ is therefore a differential domain of integrity, which can be embedded in its differential field of quotients $\mathfrak{G}'$. Since $1 \notin \Lambda$, the canonical homomorphism h of $\mathfrak{R}$ onto $\mathfrak{R}/\Lambda$ maps $\mathfrak{G}$ isomorphically; therefore we may identify each element $a \in \mathfrak{G}$ with its image $h(a) \in \mathfrak{G}'$. With this identification $\mathfrak{G}$ becomes a differential subfield of $\mathfrak{G}'$. It is now easy to see that if we set $\eta_{n\Pi j} = h(y_{n\Pi j})$ for all n, Π, j then, for each n and each $\Pi \in \mathfrak{P}_n$, $(\eta_{n\Pi 1}, \cdots, \eta_{n\Pi n})$ is a generic zero of $\Lambda(n, \Pi)$, and consequently a generic zero of Π. Therefore if we let $\mathfrak{G}^\dagger$ be an algebraic closure of $\mathfrak{G}'$ then $\mathfrak{G}^\dagger$ will have the required properties 1), 2) above. As we have seen, this suffices to prove the theorem.

Chapter II. Algebraic groups of automorphisms.

Throughout the rest of this paper $\mathcal{G}$ *will denote a differential field with algebraically closed field of constants* $\mathcal{C}$, *and* $\mathcal{F}$ *will denote a differential subfield of* $\mathcal{G}$, *with the same field of constants* $\mathcal{C}$, *such that* $\mathcal{G}$ *is finitely generated and of finite transcendence degree over* $\mathcal{F}$. *The relative algebraic closure of* $\mathcal{F}$ *in* $\mathcal{G}$ *will be denoted by* $\mathcal{F}^0$. *All differential fields mentioned will tacitly be assumed to lie in a universal extension of* $\mathcal{G}$ *fixed once and for all; in particular, every isomorphism of* $\mathcal{G}$ *will be an isomorphism into this universal extension. The identity isomorphism of* $\mathcal{G}$ *will be denoted by* ι. *The field of constants of the universal extension will be denoted by* $\mathcal{C}^*$.

1. Specializations of isomorphisms. Let $\sigma_1, \cdots, \sigma_p, \tau_1, \cdots, \tau_p$ be isomorphisms of $\mathcal{G}$; we shall say that $(\tau_1, \cdots, \tau_p)$ *is a specialization* of $\sigma_1, \cdots, \sigma_p)$ if $(\tau_i\alpha)_{1\leq i\leq p,\ \alpha\in\mathcal{C}}$ is a specialization of $(\sigma_i\alpha)_{1\leq i\leq p,\ \alpha\in\mathcal{C}}$ over $\mathcal{G}$. If $(\tau_1, \cdots, \tau_p)$ is a specialization of $(\sigma_1, \cdots, \sigma_p)$ such that $(\sigma_1, \cdots, \sigma_p)$ is a specialization of $(\tau_1, \cdots, \tau_p)$ the we shall say that $(\tau_1, \cdots, \tau_p)$ is a *generic* specialization of $(\sigma_1, \cdots, \sigma_p)$. A specialization which is not generic will be called *nongeneric*. The relation "τ is a generic specialization of σ" is an equivalence on the set of all isomorphisms of $\mathcal{G}$, and two isomorphisms of $\mathcal{G}$ which are in this relation will accordingly be called *equivalent*.

2. Isolated isomorphisms. We shall say that σ is an *isolated* isomorphism of $\mathcal{G}$ over $\mathcal{F}$ if σ is an isomorphism of $\mathcal{G}$ over $\mathcal{F}$ such that there does not exist an isomorphism of $\mathcal{G}$ over $\mathcal{F}$ of which σ is a nongeneric specialization.

Let $\eta_1, \cdots, \eta_n$ be elements such that $\mathcal{G} = \mathcal{F}\langle\eta_1, \cdots, \eta_n\rangle$. If $\sigma_1, \cdots, \sigma_p$, $\tau_1, \cdots, \tau_p$ are isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ then $(\tau_1, \cdots, \tau_p)$ is a specialization of $(\sigma_1, \cdots, \sigma_p)$ if and only if $(\tau_i\eta_j)_{1\leq i\leq p,\ 1\leq j\leq n}$ is a specialization of $(\tau_i\eta_j)_{1\leq i\leq p,\ 1\leq j\leq n}$ over $\mathcal{G}$.

Let Π be the prime differential ideal of $\mathcal{F}\{y_1, \cdots, y_n\}$ with generic zero $(\eta_1, \cdots, \eta_n)$, so that (Chapter I, Proposition 1) $\mathcal{G}\cdot\Pi$ is a perfect differential ideal of $\mathcal{G}\{y_1, \cdots, y_n\}$, and let $\Pi_1, \cdots, \Pi_h$ be the minimal prime differential ideal divisors of $\mathcal{G}\cdot\Pi$. Let $(\eta_{i1}, \cdots, \eta_{in})$ be a generic zero of Π_i; then $(\eta_{i1}, \cdots, \eta_{in})$ is also a generic zero of Π, so that there exists a unique isomorphism χ_i of $\mathcal{G}$ over $\mathcal{F}$ such that $\chi_i\eta_j = \eta_{ij}$ $(1 \leq j \leq n)$. It is obvious that if $i \neq i'$ then χ_i is not equivalent to $\chi_{i'}$. If σ is any isomorphism of $\mathcal{G}$ over $\mathcal{F}$ then $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a generic zero of Π, so that (Chapter I,

Proposition 1) $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a zero of precisely one Π_i; therefore σ is a specialization of precisely one χ_i. We have thus proved the following result.

PROPOSITION 1. *$\chi_1, \cdots, \chi_n$ are inequivalent isolated isomorphisms of $\mathfrak{G}$ over $\mathfrak{F}$, and every isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$ is a specialization of precisely one of these.*

By Proposition 1 an isomorphism σ of $\mathfrak{G}$ over $\mathfrak{F}$ is isolated if and only if σ is equivalent to χ_i for some i, that is (Chapter I, Proposition 2) if and only if

$$\begin{aligned}\partial^0\mathfrak{G}\langle\sigma\mathfrak{G}\rangle/\mathfrak{G} &= \partial^0\mathfrak{G}\langle\sigma\eta_1, \cdots, \sigma\eta_n\rangle/\mathfrak{G}\\ &= \partial^0\mathfrak{G}\langle\chi_i\eta_1, \cdots, \chi_i\eta_n\rangle/\mathfrak{G} = \operatorname{ord}\Pi_i = \operatorname{ord}\Pi = \partial^0\mathfrak{F}\langle\eta_1, \cdots, \eta_n\rangle/\mathfrak{F}\\ &= \partial^0\mathfrak{G}/\mathfrak{F}.\end{aligned}$$

Since $\mathfrak{G}\langle\sigma\mathfrak{G}\rangle = (\sigma\mathfrak{G})\langle\mathfrak{G}\rangle$ and $\partial^0\mathfrak{G}/\mathfrak{F} = \partial^0(\sigma\mathfrak{G})/\mathfrak{F}$ it follows that σ is an isolated isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$ if and only if σ^{-1} is an isolated isomorphism of $\sigma\mathfrak{G}$ over $\mathfrak{F}$. Thus we have the following result.

PROPOSITION 2. *Let σ be an isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$. The following three statements are equivalent:* 1) *σ is an isolated isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$;* 2) *σ^{-1} is an isolated isomorphism of $\sigma\mathfrak{G}$ over $\mathfrak{F}$;* 3) $\partial^0\mathfrak{G}\langle\sigma\mathfrak{G}\rangle/\mathfrak{G} = \partial^0\mathfrak{G}/\mathfrak{F}$.

Now let σ be any isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$, and suppose that σ leaves invariant some element $\zeta \in \mathfrak{G}$ which is transcendental over $\mathfrak{F}$. Let Λ be the prime differential ideal of $\mathfrak{F}\{z, y_1, \cdots, y_n\}$ with generic zero $(\zeta, \eta_1, \cdots, \eta_n)$ and let $\Lambda_1, \cdots, \Lambda_k$ be the minimal prime differential ideal divisors of the perfect differential ideal $\mathfrak{G}\cdot\Lambda$ of $\mathfrak{G}\{z, y_1, \cdots, y_n\}$. $(\zeta, \sigma\eta_1, \cdots, \sigma\eta_n)$ is a generic zero of Λ and therefore a zero of Λ_i for some i; let $(\zeta', \eta'_1, \cdots, \eta'_n)$ be a generic zero of Λ_i. Then $(\zeta', \eta'_1, \cdots, \eta'_n)$ is a generic zero of Λ, so that there exists a unique isomorphism σ' of $\mathfrak{G}$ over $\mathfrak{F}$ such that $\sigma'\zeta = \zeta'$, $\sigma'\eta_1 = \eta'_1, \cdots, \sigma'\eta_n = \eta'_n$; it is clear that σ is a specialization of σ'. If ζ were invariant under σ', that is if we had $\zeta' = \zeta$, then we would have (because ζ is transcendental over $\mathfrak{F}$)

$$\begin{aligned}&\partial^0\mathfrak{G}\langle\zeta', \eta'_1, \cdots, \eta'_n\rangle/\mathfrak{G}\\ &\quad\leq \partial^0\mathfrak{F}\langle\zeta', \eta'_1, \cdots, \eta'_n\rangle/\mathfrak{F}\langle\zeta'\rangle < \partial^0\mathfrak{F}\langle\zeta', \eta'_1, \cdots, \eta'_n\rangle/\mathfrak{F}\langle\zeta'\rangle\\ &\qquad + \partial^0\mathfrak{F}\langle\zeta'\rangle/\mathfrak{F} = \partial^0\mathfrak{F}\langle\zeta', \eta'_1, \cdots, \eta'_n\rangle/\mathfrak{F},\end{aligned}$$

or in other words $\operatorname{ord}\Lambda_i < \operatorname{ord}\Lambda$, contradicting Chapter I, Proposition 1. Therefore $\sigma'\zeta \neq \zeta$; since $\sigma\zeta = \zeta$ this means that σ is a nongeneric specialization of σ', so that σ can not be an isolated isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$.

This shows that the field of invariants of an isolated isomorphism of $\mathcal{G}$ over $\mathcal{F}$ must be contained in $\mathcal{F}^0$.

If an element of $\mathcal{G}$ is invariant under every isolated isomorphism of $\mathcal{G}$ over $\mathcal{F}$ (or, equivalently, under the isolated isomorphisms $\chi_1, \cdots, \chi_h$) then the element is invariant under every isomorphism of $\mathcal{G}$ over $\mathcal{F}$, and therefore (Kolchin [3], § 12) belongs to $\mathcal{F}$.

Consider again the prime differential ideal Π of $\mathcal{F}\{y_1, \cdots, y_n\}$ with generic zero $(\eta_1, \cdots, \eta_n)$. It is a consequence of Chapter I, Proposition 1 that the minimal prime differential ideal divisors of the perfect differential ideal $\mathcal{F}^0 \cdot \Pi$ of $\mathcal{F}^0\{y_1, \cdots, y_n\}$ are h in number, one being contained in and generating each of the minimal prime differential ideal divisors $\Pi_1, \cdots, \Pi_h$ of $\mathcal{G} \cdot \Pi$; we denote the minimal prime differential ideal divisor of $\mathcal{F}^0 \cdot \Pi$ which is contained in Π_i by Π_i^0, so that $\mathcal{G} \cdot \Pi_i^0 = \Pi_i$. Now $(\eta_1, \cdots, \eta_n)$ is a zero of precisely one Π_i^0, say of Π_1^0; because

$$\partial^0 \mathcal{F}^0\langle \eta_1, \cdots, \eta_n\rangle/\mathcal{F}^0 = \partial^0 \mathcal{F}\langle \eta_1, \cdots, \eta_n\rangle/\mathcal{F} = \operatorname{ord} \Pi = \operatorname{ord} \Pi_1$$

we see (Chapter I, Proposition 2) that $(\eta_1, \cdots, \eta_n)$ is a generic zero of Π_1^0. Also, the identity isomorphism ι is a specialization of precisely one χ_i. Because $(\eta_1, \cdots, \eta_n) = (\iota\eta_1, \cdots, \iota\eta_n)$ is a zero of Π_1^0 and $\Pi_1 = \mathcal{G} \cdot \Pi_1^0$, ι must be a specialization of χ_1. It follows from this that the restriction of χ_1 to $\mathcal{F}^0$ is the identity isomorphism of $\mathcal{F}^0$. If σ is any isomorphism of $\mathcal{G}$ such that the restriction of σ to $\mathcal{F}^0$ is the identity then $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a zero of Π_1^0 (because $(\eta_1, \cdots, \eta_n)$ is) and consequently a zero of Π_1, so that σ is a specialization of χ_1.

Collecting these remarks we have the following result.

PROPOSITION 3. *The field of invariants of an isolated isomorphism of $\mathcal{G}$ over $\mathcal{F}$ is contained in $\mathcal{F}^0$; the field of invariants of a complete set of representatives of the h equivalence classes of isolated isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ is $\mathcal{F}$. If σ_0 is an isolated isomorphism of $\mathcal{G}$ over $\mathcal{F}$ of which ι is a specialization then the field of invariants of σ_0 is $\mathcal{F}^0$; every isomorphism of $\mathcal{G}$ which leaves all the elements of $\mathcal{F}^0$ invariant is a specialization of σ_0.*

3. Strong isomorphisms. If σ is any isomorphism of $\mathcal{G}$ we denote the field of constants of $\mathcal{G}\langle\sigma\mathcal{G}\rangle$ by $\mathcal{C}_\sigma$.

We shall say that an isomorphism σ of $\mathcal{G}$ is *strong* if

$$\sigma\mathcal{G} \subset \mathcal{G}\langle\mathcal{C}^*\rangle, \quad \mathcal{G} \subset (\sigma\mathcal{G})\langle\mathcal{C}^*\rangle.$$

By Chapter I, Corollary 4 to Proposition 3 (with $\mathcal{G}$, $\mathcal{G}\langle\sigma\mathcal{G}\rangle$, $\mathcal{G}\langle\mathcal{C}^*\rangle$ playing the role of $\mathcal{E}$, $\mathcal{F}$, $\mathcal{G}$ in that Corollary) the first of these inclusions is

equivalent to $\sigma\mathfrak{G} \subset \mathfrak{G}\langle\mathfrak{k}_\sigma\rangle$; similarly, the second of these inclusions is equivalent to $\mathfrak{G} \subset (\sigma\mathfrak{G})\langle\mathfrak{k}_\sigma\rangle$. Consequently σ is strong if and only if

$$\mathfrak{G}\langle\mathfrak{k}_\sigma\rangle = \mathfrak{G}\langle\sigma\mathfrak{G}\rangle = (\sigma\mathfrak{G})\langle\mathfrak{k}_\sigma\rangle. \tag{1}$$

Obviously every automorphism of $\mathfrak{G}$ is strong.

Let $\Pi, \eta_1, \cdots, \eta_n$ have the same significance as in § 2. Because the field of constants $\mathfrak{k}$ of $\mathfrak{F}$ is algebraically closed it is easy to see that every polynomial irreducible over $\mathfrak{F}$ remains irreducible over $\mathfrak{F}\langle\mathfrak{k}^*\rangle$. It follows (Chapter I, Proposition 1) that the differential ideal $\Pi^* = \mathfrak{F}\langle\mathfrak{k}^*\rangle \cdot \Pi$ of $\mathfrak{F}\langle\mathfrak{k}^*\rangle\{y_1, \cdots, y_n\}$ is prime, and $\operatorname{ord}\Pi^* = \operatorname{ord}\Pi$.

Let σ be any strong isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$; $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a generic zero of Π, and therefore a zero of Π^*. For every finite subset $\mathfrak{c}$ of $\mathfrak{k}^*$ we have (Chapter I, Corollary 3 to Proposition 3)

$$\begin{aligned}\partial^0\mathfrak{F}\langle\mathfrak{c}\rangle\langle\sigma\eta_1, \cdots, \sigma\eta_n\rangle/\mathfrak{F}\langle\mathfrak{c}\rangle &= \partial^0(\sigma\mathfrak{G})\langle\mathfrak{c}\rangle/\mathfrak{F}\langle\mathfrak{c}\rangle \\ &= \partial^0(\sigma\mathfrak{G})\langle\mathfrak{c}\rangle/\mathfrak{F} - \partial^0\mathfrak{F}\langle\mathfrak{c}\rangle/\mathfrak{F} = \partial^0(\sigma\mathfrak{G}\langle\mathfrak{c}\rangle/\sigma\mathfrak{G} + \partial^0(\sigma\mathfrak{G})/\mathfrak{F} \\ &\quad - \partial^0\mathfrak{F}\langle\mathfrak{c}\rangle/\mathfrak{F} = \partial^0\mathfrak{k}\langle\mathfrak{c}\rangle/\mathfrak{k} + \operatorname{ord}\Pi - \partial^0\mathfrak{k}\langle\mathfrak{c}\rangle/\mathfrak{k} = \operatorname{ord}\Pi^*;\end{aligned}$$

since this holds for every finite subset $\mathfrak{c}$ of $\mathfrak{k}^*$ we infer that

$$\partial^0\mathfrak{F}\langle\mathfrak{k}^*\rangle\langle\sigma\eta_1, \cdots, \sigma\eta_n\rangle/\mathfrak{F}\langle\mathfrak{k}^*\rangle = \operatorname{ord}\Pi^*,$$

so that (Chapter I, Corollary to Proposition 2) $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a generic zero of Π^*. In the same way we also show that $(\eta_1, \cdots, \eta_n)$ is a general zero of Π^*, so that $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a generic specialization of $(\eta_1, \cdots, \eta_n)$ over $\mathfrak{F}\langle\mathfrak{k}^*\rangle$. Therefore there exists a unique isomorphism σ^* of $\mathfrak{F}\langle\mathfrak{k}^*\rangle\langle\eta_1, \cdots, \eta_n\rangle$ over $\mathfrak{F}\langle\mathfrak{k}^*\rangle$ onto $\mathfrak{F}\langle\mathfrak{k}^*\rangle\langle\sigma\eta_1, \cdots, \sigma\eta_n\rangle$ such that $\sigma^*\eta_1 = \sigma\eta_1, \cdots, \sigma^*\eta_n = \sigma\eta_n$, that is there exists a unique isomorphism σ^* of $\mathfrak{G}\langle\mathfrak{k}^*\rangle$ over $\mathfrak{F}\langle\mathfrak{k}^*\rangle$ which extends σ. Since by (1)

$$\sigma^*(\mathfrak{G}\langle\mathfrak{k}^*\rangle) = (\sigma\mathfrak{G})\langle\mathfrak{k}^*\rangle = (\sigma\mathfrak{G})\langle\mathfrak{k}_\sigma\rangle\langle\mathfrak{k}^*\rangle = \mathfrak{G}\langle\mathfrak{k}_\sigma\rangle\langle\mathfrak{k}^*\rangle = \mathfrak{G}\langle\mathfrak{k}^*\rangle,$$

we see that σ^* is an automorphism of $\mathfrak{G}\langle\mathfrak{k}^*\rangle$.

Now let us start at the other end with any automorphism σ^* of $\mathfrak{G}\langle\mathfrak{k}^*\rangle$ over $\mathfrak{F}\langle\mathfrak{k}^*\rangle$. The restriction σ of σ^* to $\mathfrak{G}$ is then an isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$. Obviously $\sigma\mathfrak{G} \subset \mathfrak{G}\langle\mathfrak{k}^*\rangle$ and $\mathfrak{G} \subset (\sigma\mathfrak{G})\langle\mathfrak{k}^*\rangle$, so that σ is strong.

We have thus proved the following result.

PROPOSITION 4. *The mapping which to each automorphism of $\mathfrak{G}\langle\mathfrak{k}^*\rangle$ over $\mathfrak{F}\langle\mathfrak{k}^*\rangle$ assigns its restriction to $\mathfrak{G}$ is one-to-one onto the set of all strong isomorphisms of $\mathfrak{G}$ over $\mathfrak{F}$.*

In virtue of Proposition 4 we may identify each strong isomorphism of $\mathcal{G}$ over $\mathcal{F}$ with the automorphism of $\mathcal{G}\langle\mathcal{L}^*\rangle$ over $\mathcal{F}\langle\mathcal{L}^*\rangle$ of which it is the restriction. This identification permits us to multiply any two strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$, and to consider the set of all of them as a group. We shall denote this group of all strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ by $\mathfrak{G}^*$, and the subgroup of $\mathfrak{G}^*$ consisting of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ by $\mathfrak{G}$.

If σ is a strong isomorphism of $\mathcal{G}$ over $\mathcal{F}$, application of σ^{-1} to (1) shows that $\mathcal{L}_\sigma \subseteq \mathcal{L}_{\sigma^{-1}}$; interchanging σ and σ^{-1} reverses the inclusion; we conclude that

$$\mathcal{L}_\sigma = \mathcal{L}_{\sigma^{-1}}. \tag{2}$$

4. Specializations of strong isomorphisms.

PROPOSITION 5. *A specialization of a strong isomorphism of $\mathcal{G}$ is always strong.*

Proof. Let σ be a strong isomorphism of $\mathcal{G}$. By (1), for each $\alpha \in \mathcal{G}$ we may write a relation $\sigma\alpha = \sum_{i=1}^{r} a_i\beta_i / \sum_{i=1}^{r} b_i\beta_i$, where a_i, b_i are constants, $\beta_1, \cdots, \beta_r$ are elements of $\mathcal{G}$ linearly independent over constants, and $\sum b_i\beta_i \neq 0$. Therefore $\beta_1, \cdots, \beta_r, \beta_1\sigma\alpha, \cdots, \beta_r\sigma\alpha$ are linearly dependent over constants so that (Chapter I, Proposition 3) the differential polynomial $W_{\theta_1 \cdots \theta_{2r}}(\beta_1, \cdots, \beta_r, \beta_1 y, \cdots, \beta_r y) \in \mathcal{G}\{y\}$ vanishes at $\sigma\alpha$ for all choices of the differential operators $\theta_1, \cdots, \theta_{2r}$ of order $< 2r$. If σ' is a specialization of σ then this differential polynomial also vanishes at $\sigma'\alpha$ so that (Chapter I, Proposition 3) there exist elements $a'_1, \cdots, a'_r, b'_1, \cdots, b'_r \in \mathcal{L}_{\sigma'}$ not all 0 such that $\sum a'_i\beta_i - \sum b'_i\beta_i\sigma'\alpha = 0$. Since $\beta_1, \cdots, \beta_r$ are linearly independent over constants not every b'_i is 0 so that $\sum b'_i\beta_i \neq 0$ and

$$\sigma'\alpha = \sum a'_i\beta_i / \sum b'_i\beta_i \in \mathcal{G}\langle\mathcal{L}_{\sigma'}\rangle.$$

Thus $\mathcal{G}\langle\sigma'\mathcal{G}\rangle = \mathcal{G}\langle\mathcal{L}_{\sigma'}\rangle$.

Again by (1), for each $\alpha \in \mathcal{G}$ we may write $\alpha = \sum_{i=1}^{r} a_i\sigma\beta_i / \sum_{i=1}^{r} b_i\sigma\beta_i$, where a_i, b_i are constants, $\sigma\beta_1, \cdots, \sigma\beta_r$ are elements of $\sigma\mathcal{G}$ linearly independent over constants, and $\sum b_i\sigma\beta_i \neq 0$ (so that $\beta_1, \cdots, \beta_r$ are elements of $\mathcal{G}$ linearly independent over constants and $\sum b_i\beta_i \neq 0$). Therefore the differential polynomial $W_{\theta_1 \cdots \theta_{2r}}(y_1, \cdots, y_r, \alpha y_1, \cdots, \alpha y_r) \in \mathcal{G}\{y_1, \cdots, y_r\}$ vanishes at $(\sigma\beta_1, \cdots, \sigma\beta_r)$ for every choice of $\theta_1, \cdots, \theta_{2r}$ of order $< 2r$, so that this differential polynomial also vanishes at $(\sigma'\beta_1, \cdots, \sigma'\beta_r)$. This implies that there exist constants $a'_i, b'_i \in \mathcal{L}_{\sigma'}$ not all 0 such that $\sum a'_i\sigma'\beta_i - \sum \alpha b'_i\sigma'\beta_i = 0$;

because $\beta_1, \cdots, \beta_r$ are linearly independent over constants $\sigma'\beta_1, \cdots, \sigma'\beta_r$ are too, so that $\sum b'_i\sigma'\beta_i \neq 0$. Therefore

$$\alpha = \sum a'_i\sigma'\beta_i / \sum b'_i\sigma'\beta_i \in (\sigma'\mathcal{G})\langle\mathcal{C}_{\sigma'}\rangle,$$

so that $\mathcal{G}\langle\sigma'\mathcal{G}\rangle = (\sigma'\mathcal{G})\langle\mathcal{C}_{\sigma'}\rangle$. It follows that σ' satisfies (1) and σ' is strong.

REMARK. We observe from the proof of Proposition 5 that if σ is an isomorphism of $\mathcal{G}$ which satisfies the first (second) equation (1) then every specialization of σ also satisfies the first (second) equation (1).

PROPOSITION 6. *If $\sigma_1, \cdots, \sigma_p$ are strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ and if $(\tau_1, \cdots, \tau_p)$ is a specialization of $(\sigma_1, \cdots, \sigma_p)$ then $(\tau_1^{-1}, \tau_1^{-1}\tau_2, \cdots, \tau_1^{-1}\tau_p)$ is a specialization of $(\sigma_1^{-1}, \sigma_1^{-1}\sigma_2, \cdots, \sigma_1^{-1}\sigma_p)$.*

Proof. Let $F \in \mathcal{G}\{(z_{1j})_{1\leq j\leq q}, (z_{ij})_{2\leq i\leq p,\ 1\leq j\leq q}\}$; if we denote the coefficients in F by $\beta_1, \cdots, \beta_r$ then we may write

$$F((z_{1j})_{1\leq j\leq q}, (z_{ij})_{2\leq i\leq p,\ 1\leq j\leq q}) = G((z_{1j})_{1\leq j\leq q}, (z_{ij})_{2\leq i\leq p,\ 1\leq j\leq q}, (\beta_k)_{1\leq k\leq r}),$$

where $G \in \mathcal{F}\{(z_{1j})_{1\leq j\leq q}, (z_{ij})_{2\leq i\leq p,\ 1\leq j\leq q}, (y_k)_{1\leq k\leq r}\}$. If F vanishes at $((\sigma_1^{-1}\alpha_j)_{1\leq j\leq q}, (\sigma_1^{-1}\sigma_i\alpha_j)_{2\leq i\leq p,\ 1\leq j\leq q})$, that is if G vanishes at

$$((\sigma_1^{-1}\alpha_j)_{1\leq j\leq q}, (\sigma_1^{-1}\sigma_i\alpha_j)_{2\leq i\leq p,\ 1\leq j\leq q}, (\beta_k)_{1\leq k\leq r}),$$

then application of σ_1 shows that G vanishes at

$$((\alpha_j)_{1\leq j\leq q}, (\sigma_i\alpha_j)_{2\leq i\leq p,\ 1\leq j\leq q}, (\sigma_1\beta_k)_{1\leq k\leq r});$$

since $(\tau_1, \cdots, \tau_p)$ is a specialization of $(\sigma_1, \cdots, \sigma_p)$ this implies that G vanishes at

$$((\alpha_j)_{1\leq j\leq q}, (\tau_i\alpha_j)_{2\leq i\leq p,\ 1\leq j\leq q}, (\tau_1\beta_k)_{1\leq k\leq r});$$

application of τ_1^{-1} shows that G vanishes at

$$((\tau_1^{-1}\alpha_j)_{1\leq j\leq q}, (\tau_1^{-1}\tau_i\alpha_j)_{2\leq i\leq p,\ 1\leq j\leq q}, (\beta_k)_{1\leq k\leq r}),$$

that is that F vanishes at $((\tau_1^{-1}\alpha_j)_{1\leq j\leq q}, (\tau_1^{-1}\tau_i\alpha_j)_{2\leq i\leq p,\ 1\leq j\leq q})$. It follows that $\tau_1^{-1}, \tau_1^{-1}\tau_2, \cdots, \tau_1^{-1}\tau_p)$ is a specialization of $(\sigma_1^{-1}, \sigma_1^{-1}\sigma_2, \cdots, \sigma_1^{-1}\sigma_p)$.

Let $\sigma_1, \cdots, \sigma_p$ be isomorphisms of $\mathcal{G}$ which have the following property: whenever $\tau_1, \cdots, \tau_p$ are isomorphisms of $\mathcal{G}$ such that τ_i is a specialization of σ_i $(1 \leq i \leq p)$ then $(\tau_1, \cdots, \tau_p)$ is a specialization of $(\sigma_1, \cdots, \sigma_p)$; we shall say under these circumstances that $\sigma_1, \cdots, \sigma_p$ are *independent*.

PROPOSITION 7. *The strong isomorphisms* $\sigma_1, \cdots, \sigma_p$ *of* $\mathfrak{G}$ *over* $\mathfrak{F}$ *are independent if and only if* $\partial^0 \mathfrak{G}\langle \sigma_1 \mathfrak{G}, \cdots, \sigma_p \mathfrak{G}\rangle/\mathfrak{G} = \sum_{i=1}^{p} \partial^0 \mathfrak{G}\langle \sigma_i \mathfrak{G}\rangle/\mathfrak{G}$.

Proof. Let Λ_i denote the prime differential ideal in $\mathfrak{G}\{y_{i1}, \cdots, y_{im}\}$ with generic zero $(\sigma_i\eta_i, \cdots, \sigma_i\eta_n)$, where as before $\eta_1, \cdots, \eta_n$ generate $\mathfrak{G} : \mathfrak{G} = \mathfrak{F}\langle \eta_1, \cdots, \eta_n\rangle$. If $(\eta'_{i1}, \cdots, \eta'_{in})$ is a generic zero of Λ_i then $(\eta'_{i1}, \cdots, \eta'_{in})$ is a generic specialization of $(\sigma_i\eta_1, \cdots, \sigma_i\eta_n)$ over $\mathfrak{G}$, so that there exists an isomorphism of $\mathfrak{G}\langle \sigma_i\mathfrak{G}\rangle = \mathfrak{G}\langle \sigma_i\eta_1, \cdots, \sigma_i\eta_n\rangle$ onto $\mathfrak{G}\langle \eta'_{i1}, \cdots, \eta'_{in}\rangle$ over $\mathfrak{G}$; since $\mathfrak{G}\langle \sigma_i\mathfrak{G}\rangle = \mathfrak{G}\langle \mathfrak{C}_{\sigma_i}\rangle$ we have $\mathfrak{G}\langle \eta'_{i1}, \cdots, \eta'_{in}\rangle = \mathfrak{G}\langle \mathfrak{C}'_i\rangle$, where $\mathfrak{C}'_i$ is the field of constants of $\mathfrak{G}\langle \eta'_{i1}, \cdots, \eta'_{in}\rangle$. The field of constants of $\mathfrak{G}$ is $\mathfrak{C}$, which is algebraically closed; it easily follows that every polynomial which is irreducible over $\mathfrak{G}$ remains irreducible over

$$\mathfrak{G}_{i_0} = \mathfrak{G}\langle \mathfrak{C}'_i, \cdots, \mathfrak{C}'_{i_0-1}, \mathfrak{C}'_{i_0+1}, \cdots, \mathfrak{C}'_p\rangle = \mathfrak{G}\langle (\eta'_{ij})_{1\leq i\leq p,\, i\neq i_0,\, 1\leq j\leq n}\rangle,$$

so that (Chapter I, Proposition 1) $\mathfrak{G}_{i_0} \cdot \Lambda_{i_0}$ is prime $(1 \leq i_0 \leq p)$. It follows (see remark following the proof of Proposition 4 of Chapter I) that the ideal Λ generated by $\bigcup_{i=1}^{p} \Lambda_i$ in $\mathfrak{G}\{(y_{ij})_{1\leq i\leq p, 1\leq j\leq n}\}$ is a prime differential ideal. Since the order of each Λ_i is finite so is the order of Λ, and ord $\Lambda = \sum$ ord Λ_i. Therefore (Chapter I, Corollary to Proposition 2) $(\sigma_i\eta_j)_{1\leq i\leq p,\, 1\leq j\leq n}$ is a generic zero of Λ if and only if

$$\partial^0 \mathfrak{G}\langle \sigma_1\mathfrak{G}, \cdots, \sigma_p\mathfrak{G}\rangle/\mathfrak{G} = \partial^0 \mathfrak{G}\langle (\sigma_i\eta_j)_{1\leq i\leq p,\, 1\leq j\leq n}\rangle/\mathfrak{G} = \text{ord } \Lambda$$
$$= \sum \text{ord } \Lambda_i = \sum \partial^0 \mathfrak{G}\langle \sigma_i\eta_1, \cdots, \sigma_i\eta_n\rangle/\mathfrak{G} = \partial^0 \mathfrak{G}\langle \sigma_i\mathfrak{G}\rangle/\mathfrak{G}.$$

It is easy to see, however, that $\sigma_1, \cdots, \sigma_p$ are independent if and only if $(\sigma_i\eta_j)_{1\leq i\leq p,\, 1\leq j\leq n}$ is a generic zero of Λ. This completes the proof.

PROPOSITION 8. *If* $\sigma_1, \cdots, \sigma_p$ *are independent strong isomorphisms of* $\mathfrak{G}$ *over* $\mathfrak{F}$ *and if* $\tau_1, \cdots, \tau_p$ *are isomorphisms of* $\mathfrak{G}$ *such that* τ_i *is a specialization of* σ_i $(1 \leq i \leq p)$ *then* $(\tau_1, \tau_1\tau_2, \cdots, \tau_1\tau_p)$ *is a specialization of* $(\sigma_1, \sigma_1\sigma_2, \cdots, \sigma_1\sigma_p)$.

Proof. By (1) and (2) we have

$$\partial^0 \mathfrak{G}\langle \sigma_1^{-1}\mathfrak{G}, \sigma_2\mathfrak{G}, \cdots, \sigma_p\mathfrak{G}\rangle/\mathfrak{G} = \partial^0 \mathfrak{G}\langle \mathfrak{C}_{\sigma_1}, \mathfrak{C}_{\sigma_2}, \cdots, \mathfrak{C}_{\sigma_p}\rangle\mathfrak{G}$$
$$= \partial^0 \mathfrak{G}\langle \sigma_1\mathfrak{G}, \sigma_2\mathfrak{G}, \cdots, \sigma_p\mathfrak{G}\rangle/\mathfrak{G},$$

which (Proposition 7) equals

$$\textstyle\sum_{1\leq i\leq p} \partial^0 \mathfrak{G}\langle \sigma_i\mathfrak{G}\rangle/\mathfrak{G} = \partial^0 \mathfrak{G}\langle \sigma_1^{-1}\mathfrak{G}\rangle/\mathfrak{G} + \sum_{2\leq i\leq p} \partial^0 \mathfrak{G}\langle \sigma_i\mathfrak{G}\rangle/\mathfrak{G}.$$

Therefore (Proposition 7) $\sigma_1^{-1}, \sigma_2, \cdots, \sigma_p$ are independent. If τ_i is a specialization of σ_i $(1 \leq i \leq p)$ then (Proposition 6) τ_1^{-1} is a specialization of σ_1^{-1}, whence $(\tau_1^{-1}, \tau_2, \cdots, \tau_p)$ is a specialization of $(\sigma_1^{-1}, \sigma_2, \cdots, \sigma_p)$. By Proposition 6 it follows that $(\tau_1, \tau_1\tau_2, \cdots, \tau_1\tau_p)$ is a specialization of $(\sigma_1, \sigma_1\sigma_2, \cdots, \sigma_1\sigma_p)$.

If $W(X_1, \cdots, X_p)$ is a word in $X_1, \cdots, X_p$, that is if it is an element of the free group generated by $X_1, \cdots, X_p$, and if $\sigma_1, \cdots, \sigma_p$ are strong isomorphisms of $\mathfrak{G}$ over $\mathfrak{F}$, then $W(\sigma_1, \cdots, \sigma_p)$, the meaning of which is obvious, is itself a strong isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$.

If σ is a strong isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$ then, since $\mathfrak{G}$ is finitely generated over $\mathfrak{F}$, $\sigma\mathfrak{G}$ is too, so that $\mathfrak{G}\langle\sigma\mathfrak{G}\rangle = \mathfrak{G}\langle\mathfrak{C}_\sigma\rangle$ is finitely generated over $\mathfrak{G}$; using Chapter I, § 4, it is not difficult to see that then $\mathfrak{C}_\sigma$ is finitely generated over $\mathfrak{C}$.

PROPOSITION 9. *Let $\sigma_1, \cdots, \sigma_p$ be strong isomorphisms of $\mathfrak{G}$ over $\mathfrak{F}$, let $\gamma_{i1}, \cdots, \gamma_{iq_i}$ be constants such that $\mathfrak{C}_{\sigma_i} = \mathfrak{C}\langle\gamma_{i1}, \cdots, \gamma_{iq_i}\rangle$ $(1 \leq i \leq p)$, let $W_1(X_1, \cdots, X_p), \cdots, W_r(X_1, \cdots, X_p)$ be words, let $\xi_1, \cdots, \xi_r$ be elements of $\mathfrak{G}$, and let N be a differential polynomial in $\mathfrak{G}\{w_1, \cdots, w_r\}$ which does not vanish at $(W_1(\sigma_1, \cdots, \sigma_p)\xi_1, \cdots, W_r(\sigma_1, \cdots, \sigma_p)\xi_r)$. Then there exists a polynomial $M \in \mathfrak{C}[(u_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}]$ which does not vanish at $(\gamma_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}$ and which has the following property: if $(c_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}$ is a family of constants which is a specialization of $(\gamma_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}$ over $\mathfrak{C}$ which does not annul M then for each i $(1 \leq i \leq p)$ there exists a unique isomorphism τ_i of $\mathfrak{G}$ such that $((\tau_i\alpha)_{\alpha\in\mathfrak{G}}, (c_{ik})_{1\leq k\leq q_i})$ is a specialization of $((\sigma_i\alpha)_{\alpha\in\mathfrak{G}}, (\gamma_{ik})_{1\leq k\leq q_i})$ over $\mathfrak{G}$, $\mathfrak{C}_{\tau_i} = \mathfrak{C}\langle c_{i1}, \cdots, c_{iq_i}\rangle$, N does not vanish at*

$$(W_1(\tau_1, \cdots, \tau_p)\xi_1, \cdots, W_r(\tau_1, \cdots, \tau_p)\xi_r),$$

and, for every finite family

$$(W'_1(X_1, \cdots, X_p), \cdots, W'_s(X_1, \cdots, X_p))$$

of words,

$$(W'_1(\tau_1, \cdots, \tau_p), \cdots, W'_s(\tau_1, \cdots, \tau_p))$$

is a specialization of

$$(W'_1(\sigma_1, \cdots, \sigma_p), \cdots, W'_s(\sigma_1, \cdots, \sigma_p)).$$

Proof. By (1) there exist

$$A_{ij}, B_{ij} \in \mathfrak{F}\{y_1, \cdots, y_n\}[u_{i1}, \cdots, u_{iq_i}] \qquad (1 \leq i \leq p, 1 \leq j \leq n),$$

$$C_{ik}, D_{ik} \in \mathfrak{F}\{y_1, \cdots, y_n, z_{i1}, \cdots, z_{in}\} \qquad (1 \leq i \leq p, 1 \leq k \leq q_i),$$

$$E_{ij}, F_{ij} \in \mathfrak{F}\{z_{i1}, \cdots, z_{in}\}[u_{i1}, \cdots, u_{iq_i}] \qquad (1 \leq i \leq p, 1 \leq j \leq n)$$

such that, for all i, j, k,

$$B_{ij}(\eta_1, \cdots, \eta_n, \gamma_{i1}, \cdots, \gamma_{iq_i}) \neq 0,$$

$$D_{ik}(\eta_1, \cdots, \eta_n, \sigma_i\eta_1, \cdots, \sigma_i\eta_n) \neq 0,$$

$$F_{ij}(\sigma_i\eta_1, \cdots, \sigma_i\eta_n, \gamma_{i1}, \cdots, \gamma_{iq_i}) \neq 0,$$

$$(3) \quad \sigma_i\eta_j = A_{ij}(\eta_1, \cdots, \eta_n, \gamma_{i1}, \cdots, \gamma_{iq_i})/B_{ij}(\eta_1, \cdots, \eta_n, \gamma_{i1}, \cdots, \gamma_{iq_i}),$$

$$\gamma_{ik} = C_{ik}(\eta_1, \cdots, \eta_n, \sigma_i\eta_1, \cdots, \sigma_i\eta_n)/D_{ik}(\eta_1, \cdots, \eta_n, \sigma_i\eta_1, \cdots, \sigma_i\eta_n),$$

$$(4) \quad \eta_j = E_{ij}(\sigma_i\eta_1, \cdots, \sigma_i\eta_n, \gamma_{i1}, \cdots, \gamma_{iq_i})/F_{ij}(\sigma_i\eta_1, \cdots, \sigma_i\eta_n, \gamma_{i1}, \cdots, \gamma_{iq_i}).$$

Let $(c_{ik})_{1\leq i\leq p,\ 1\leq k\leq q_i}$ be a family of constants which is a specialization of $(\gamma_{ik})_{1\leq i\leq p,\ 1\leq k\leq q_i}$ over $\mathscr{C}$, and therefore (Chapter I, Corollary 2 to Proposition 3) over $\mathscr{G}$. If

$$(5) \quad \prod_{i,j} B_{ij}(\eta_1, \cdots, \eta_n, c_{i1}, \cdots, c_{iq_i}) \neq 0$$

and if we set

$$(6) \quad \zeta_{ij} = A_{ij}(\eta_1, \cdots, \eta_n, c_{i1}, \cdots, c_{iq_i})/B_{ij}(\eta_1, \cdots, \eta_n, c_{i1}, \cdots, c_{iq_i})$$

then $((\zeta_{ij})_{1\leq i\leq p,\ 1\leq j\leq n},\ (c_{ik})_{1\leq i\leq p,\ 1\leq k\leq q_i})$ is a specialization of

$$((\sigma_i\eta_j)_{1\leq i\leq p,\ 1\leq j\leq n},\ (\gamma_{ik})_{1\leq i\leq p,\ 1\leq k\leq q_i})$$

over $\mathscr{G}$. If moreover

$$(7) \quad \prod_{i,k} D_{ik}(\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in}) \neq 0$$

then

$$(8) \quad c_{ik} = C_{ik}(\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in})/D_{ik}(\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in}).$$

If in addition

$$(9) \quad \prod_{i,j} F_{ij}(\zeta_{i1}, \cdots, \zeta_{in}, c_{i1}, \cdots, c_{iq_i}) \neq 0$$

then

$$(10) \quad \eta_j = E_{ij}(\zeta_{i1}, \cdots, \zeta_{in}, c_{i1}, \cdots, c_{iq_i})/F_{ij}(\zeta_{i1}, \cdots, \zeta_{in}, c_{i1}, \cdots, c_{iq_i}).$$

Assuming that (5), (7), (9) hold we see that $(\zeta_{i1}, \cdots, \zeta_{in})$ is a specialization of $(\sigma_i\eta_1, \cdots, \sigma_i\eta_n)$ over $\mathscr{G}$, and therefore of $(\eta_1, \cdots, \eta_n)$ over $\mathscr{F}$, such that

$$\partial^0\mathscr{F}\langle\zeta_{i1}, \cdots, \zeta_{in}\rangle/\mathscr{F} - \partial^0\mathscr{F}\langle\eta_1, \cdots, \eta_n\rangle/\mathscr{F}$$

$$= \partial^0\mathscr{F}\langle\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in}\rangle/\mathscr{F} - \partial^0\mathscr{F}\langle\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in}\rangle/\mathscr{F}\langle\zeta_{i1}, \cdots$$

$$- \partial^0\mathscr{F}\langle\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in}\rangle/\mathscr{F}$$

$$+ \partial^0\mathscr{F}\langle\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in}\rangle/\mathscr{F}\langle\eta_1, \cdots, \eta_n\rangle,$$

which by (6), (8), (10) equals

$$-\partial^0\mathfrak{F}\langle\zeta_{i1},\cdots,\zeta_{in},c_{i1},\cdots,c_{iq_i}\rangle/\mathfrak{F}\langle\zeta_{i1},\cdots,\zeta_{in}\rangle$$
$$+\partial^0\mathfrak{F}\langle\eta_1,\cdots,\eta_n,c_{i1},\cdots,c_{iq_i}\rangle/\mathfrak{F}\langle\eta_1,\cdots,\eta_n\rangle,$$

which by Chapter I, Corollary 3 to Proposition 3, is ≥ 0. It follows (Chapter I, Proposition 2) that $(\zeta_{i1},\cdots,\zeta_{in})$ is a generic specialization of $(\eta_1,\cdots,\eta_n)$ over $\mathfrak{F}$, so that there exists a unique isomorphism τ_i of $\mathfrak{G}=\mathfrak{F}\langle\eta_1,\cdots,\eta_n\rangle$ over $\mathfrak{F}$ such that $\tau_i\eta_j=\zeta_{ij}$ $(1\leq j\leq n)$. By (6), (8), (10)

$$\mathfrak{G}\langle\tau_i\mathfrak{G}\rangle=\mathfrak{G}\langle c_{i1},\cdots,c_{iq_i}\rangle=(\tau_i\mathfrak{G})\langle c_{i1},\cdots,c_{iq_i}\rangle,$$

so that $\mathfrak{C}_{\tau_i}=\mathfrak{C}\langle c_{i1},\cdots,c_{iq_i}\rangle$. It is apparent that τ_i is the unique isomorphism of $\mathfrak{G}$ such that $((\tau_i\alpha)_{\alpha\in\mathfrak{G}},(c_{ik})_{1\leq k\leq q_i})$ is a specialization of $((\sigma_i\alpha)_{\alpha\in\mathfrak{G}},(\gamma_{ik})_{1\leq k\leq q_i})$ over $\mathfrak{G}$.

By (4) and (10)

(11) $\sigma_i^{-1}\eta_j=E_{ij}(\eta_1,\cdots,\eta_n,\gamma_{i1},\cdots,\gamma_{iq_i})/F_{ij}(\eta_1,\cdots,\eta_n,\gamma_{i1},\cdots,\gamma_{iq_i})$,

(12) $\tau_i^{-1}\eta_j=E_{ij}(\eta_1,\cdots,\eta_n,c_{i1},\cdots,c_{iq_i})/F_{ij}(\eta_1,\cdots,\eta_n,c_{i1},\cdots,c_{iq_i})$.

From (3), (6), (11), (12) it is not difficult to see that for every word $W(X_1,\cdots,X_p)$ there exist

$$G_j^W,H_j^W\in\mathfrak{F}\{y_1,\cdots,y_n\}[(u_{ik})_{1\leq i\leq p,\,1\leq k\leq q_i}]$$

such that

$$H_j^W(\eta_1,\cdots,\eta_n,(\gamma_{ik}))\neq 0,\qquad H_j^W(\eta_1,\cdots,\eta_n,(c_{ik}))\neq 0$$

and

$$W(\sigma_1,\cdots,\sigma_p)\eta_j=G_j^W(\eta_1,\cdots,\eta_n,(\gamma_{ik}))/H_j^W(\eta_1,\cdots,\eta_n,(\gamma_{ik})),$$

$$W(\tau_1,\cdots,\tau_p)\eta_j=G_j^W(\eta_1,\cdots,\eta_n,(c_{ik}))/H_j^W(\eta_1,\cdots,\eta_n,(c_{ik})).$$

It follows, for any finite family of words

$$W'_1(X_1,\cdots,X_p),\cdots,W'_s(X_1,\cdots,X_p),$$

that $(W'_l(\tau_1,\cdots,\tau_p)\eta_j)_{1\leq l\leq s,\,1\leq j\leq n}$ is a specialization of

$$W'_l(\sigma_1,\cdots,\sigma_p)\eta_j)_{1\leq l\leq s,\,1\leq j\leq n}$$

over $\mathfrak{G}$, so that

$$(W'_1(\tau_1,\cdots,\tau_p),\cdots,W'_s(\tau_1,\cdots,\tau_p))$$

is a specialization of

$$(W'_1(\sigma_1,\cdots,\sigma_p),\cdots,W'_s(\sigma_1,\cdots,\sigma_p)).$$

It also follows that for each h $(1\leq h\leq r)$ there exist

$$I_h,J_h\in\mathfrak{F}\{y_1,\cdots,y_n\}[(u_{ik})_{1\leq i\leq p,\,1\leq k\leq q_i}]$$

with

$$J_h(\eta_1, \cdots, \eta_n, (\gamma_{ik})) \neq 0, \qquad J_h(\eta_1, \cdots, \eta_n, (c_{ik})) \neq 0$$

such that

$$W_h(\sigma_1, \cdots, \sigma_p)\xi_h = I_h(\eta_1, \cdots, \eta_n, (\gamma_{ik}))/J_h(\eta_1, \cdots, \eta_n, (\gamma_{ik})),$$

$$W_h(\tau_1, \cdots, \tau_p)\xi_h = I_h(\eta_1, \cdots, \eta_n, (c_{ik}))/J_h(\eta_1, \cdots, \eta_n, (c_{ik})),$$

and therefore that there exist

$$U, V \in \mathfrak{F}\{y_1, \cdots, y_n\}[(u_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}]$$

such that

$$V(\eta_1, \cdots, \eta_n), (\gamma_{ik})) \neq 0, \qquad V(\eta_1, \cdots, \eta_n(c_{ik})) \neq 0$$

and

$$N(W_1(\sigma_1, \cdots, \sigma_p)\xi_1, \cdots, W_r(\sigma_1, \cdots, \sigma_p)\xi_r) = U(\eta_1, \cdots, \eta_n, (\gamma_{ik}))/V(\eta_1, \cdots, \eta_n, (\gamma_{ik})),$$

$$N(W_1(\tau_1, \cdots, \tau_p)\xi_1, \cdots, W_r(\tau_1, \cdots, \tau_p)\xi_r) = U(\eta_1, \cdots, \eta_n, (c_{ik}))/V(\eta_1, \cdots, \eta_n, (c_{ik})),$$

Now there exists a polynomial $M' \in \mathfrak{G}[(u_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}]$ which has the two properties that it does not vanish at $(\gamma_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}$ and that if it does not vanish at $(c_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}$ then (5), (7), (9) hold and $U(\eta_1, \cdots, \eta_n, (c_{ik})) \neq 0$. Therefore (Chapter I, Corollary 2 to Proposition 3) there exists a polynomial $M \in \mathfrak{G}[(u_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}]$ with the same two properties. This M has the property described in the statement of the proposition.

COROLLARY 1. *Let $\eta_1, \cdots, \eta_n$ be elements of $\mathfrak{G}$ such that $\mathfrak{G} = \mathfrak{F}\langle\eta_1, \cdots, \eta_n\rangle$, let q, r be integers such that $1 \leq q \leq r$, let $W_1(X_1, \cdots, X_p), \cdots, W_r(X_1, \cdots, X_p)$ be words, let $\sigma_1, \cdots, \sigma_p$ be strong isomorphisms of $\mathfrak{G}$ over $\mathfrak{F}$, and let $Q \in \mathfrak{G}\{(y_{ij})_{1\leq i\leq r,\, 1\leq j\leq n}\}$ not vanish at $(W_i(\sigma_1, \cdots, \sigma_p)\eta_j)_{1\leq i\leq r,\, 1\leq j\leq n}$. Then there exists a $P \in \mathfrak{G}\{(y_{ij})_{1\leq i\leq q,\, 1\leq j\leq n}\}$ which does not vanish at $(W_i(\sigma_1, \cdots, \sigma_p)\eta_j)_{1\leq i\leq q,\, 1\leq j\leq n}$ such that for every specialization $(\rho'_1, \cdots, \rho'_q)$ of $(W_1(\sigma_1, \cdots, \sigma_p), \cdots, W_q(\sigma_1, \cdots, \sigma_p))$ for which $P((\rho'_i\eta_j)_{1\leq i\leq q,\, 1\leq j\leq n}) \neq 0$ there exists strong isomorphisms $\sigma'_1, \cdots, \sigma'_p$ of G such that $\rho'_i = W_i(\sigma'_1, \cdots, \sigma'_p)$ $(1 \leq i \leq q)$,*

$$(\sigma'_1, \cdots, \sigma'_p, W_1(\sigma'_1, \cdots, \sigma'_p), \cdots, W_r(\sigma'_1, \cdots, \sigma'_p))$$

is a specialization of

$$(\sigma_1, \cdots, \sigma_p, W_1(\sigma_1, \cdots, \sigma_p), \cdots, W_r(\sigma_1, \cdots, \sigma_p)),$$

and $Q((W_i(\sigma'_1, \cdots, \sigma'_p)\eta_j)_{1\leq i\leq r,\, 1\leq j\leq n}) \neq 0$.

COROLLARY 2. *Let σ be a strong isomorphism of $\mathcal{G}$ over $\mathcal{F}$, let $\xi_1, \cdots, \xi_s \in \mathcal{G}$, let $N \in \mathcal{G}\{w_1, \cdots, w_s\}$ not vanish at $(\sigma\xi_1, \cdots, \xi\sigma_s)$. Then there exists an automorphism τ of $\mathcal{G}$ over $\mathcal{F}$ such that τ is a specialization of σ and $N(\tau\xi_1, \cdots, \tau\xi_s) \neq 0$.*

COROLLARY 3. *Let σ be a strong isomorphism of $\mathcal{G}$ over $\mathcal{F}$, let $\mathcal{F}_1$ be a differential field between $\mathcal{F}$ and $\mathcal{G}$, and suppose that the restriction σ' of σ to $\mathcal{F}_1$ is a strong isomorphism of $\mathcal{F}_1$. Then there exist a finite number of nongeneric specializations $\sigma'_1, \cdots, \sigma'_s$ of σ' such that every specialization τ' of σ' which is not a specialization of σ'_j $(1 \leq j \leq s)$ is the restriction to $\mathcal{F}_1$ of some specialization τ of σ.*

Proof. Let $\gamma_1, \cdots, \gamma_p, \gamma_{p+1}, \cdots, \gamma_q$ be constants such that $\mathcal{F}_1\langle\sigma'\mathcal{F}_1\rangle = \mathcal{F}_1\langle\gamma_1, \cdots, \gamma_p\rangle$, $\mathcal{G}\langle\sigma\mathcal{G}\rangle = \mathcal{G}\langle\gamma_1, \cdots, \gamma_q\rangle$. By Proposition 9 there exists a polynomial $M \in \mathcal{C}[u_1, \cdots, u_q]$ with $M(\gamma_1, \cdots, \gamma_q) \neq 0$ such that whenever $c_1, \cdots, c_q$ are constants such that $(c_1, \cdots, c_q)$ is a specialization of $(\gamma_1, \cdots, \gamma_q)$ over $\mathcal{C}$ with $M(c_1, \cdots, c_n) \neq 0$ then there exist an isomorphism τ of $\mathcal{G}$ such that $((\tau\alpha)_{\alpha\in\mathcal{G}}, c_1, \cdots, c_q)$ is a specialization of $((\sigma\alpha)_{\alpha\in\mathcal{G}}, \gamma_1, \cdots, \gamma_q)$ over $\mathcal{G}$, and a unique isomorphism τ' of $\mathcal{F}_1$ such that $((\tau'\alpha)_{\alpha\in\mathcal{F}_1}, c_1, \cdots, c_p)$ is a specialization of $((\sigma'\alpha)_{\alpha\in\mathcal{F}_1}, \gamma_1, \cdots, \gamma_p)$ over $\mathcal{F}_1$. There also exists a polynomial $K \in \mathcal{C}[u_1, \cdots, u_p]$ with $K(\gamma_1, \cdots, \gamma_p) \neq 0$ such that every specialization $(c_1, \cdots, c_p)$ of $(\gamma_1, \cdots, \gamma_p)$ with $K(c_1, \cdots, c_p) \neq 0$ can be extended to a specialization $(c_1, \cdots, c_q)$ of $(\gamma_1, \cdots, \gamma_q)$ over $\mathcal{C}$ with $M(c_1, \cdots, c_q) \neq 0$. We now write, for $1 \leq i \leq p$,

$$\gamma_i = R_i(\sigma'\theta_1, \cdots, \sigma'\theta_l)/S(\sigma'\theta_1, \cdots, \sigma'\theta_l),$$

where $\theta_1, \cdots, \theta_l$ are elements of $\mathcal{F}_1$, such that $\mathcal{F}_1 = \mathcal{F}\langle\theta_1, \cdots, \theta_l\rangle$, R_i and $S \in \mathcal{F}_1\{y_1, \cdots, y_l\}$, and $S(\sigma'\theta_1, \cdots, \sigma'\theta_l) \neq 0$. If τ' is any specialization of σ' such that $S(\tau'\theta_1, \cdots, \tau'\theta_l) \neq 0$ and if we set

$$c_i = R_i(\tau'\theta_1, \cdots, \tau'\theta_l)/S(\tau'\theta_1, \cdots, \tau'\theta_l),$$

then $c_1, \cdots, c_p$ are constants such that $(c_1, \cdots, c_p)$ is a specialization of $(\gamma_1, \cdots, \gamma_p)$ over $\mathcal{C}$. If in addition

$$K\left(\frac{R_1(\tau'\theta_1, \cdots, \tau'\theta_l)}{S(\tau'\theta_1, \cdots, \tau'\theta_l)}, \cdots, \frac{R_p(\tau'\theta_1, \cdots, \tau'\theta_l)}{S(\tau'\theta_1, \cdots, \tau'\theta_l)}\right) \neq 0,$$

that is $K(c_1, \cdots, c_p) \neq 0$, then we may extend $(c_1, \cdots, c_p)$ to a specialization $(c_1, \cdots, c_q)$ of $(\gamma_1, \cdots, \gamma_q)$ over $\mathcal{C}$ such that $M(c_1, \cdots, c_q) \neq 0$, so that there exists an isomorphism τ of $\mathcal{G}$ such that $((\tau\alpha)_{\alpha\in\mathcal{G}}, c_1, \cdots, c_q)$ is a specialization of $((\sigma\alpha)_{\alpha\in\mathcal{G}}, \gamma_1, \cdots, \gamma_q)$ over $\mathcal{G}$. Under these circum-

stances τ will be a specialization of σ, and τ' will be the restriction of τ to $\mathfrak{F}_1$. Thus, the specializations τ' of σ' which can *not* be extended to a specialization of σ have the property that $(\tau'\theta_1, \cdots, \tau'\theta_l)$ is a zero of SL, where L is a differential polynomial in $\mathfrak{F}_1\{y_1, \cdots, y_l\}$ obtained by multiplying $K(S^{-1}R_1, \cdots, S^{-1}R_l)$ by a power of S; of course $(\sigma'\theta_1, \cdots, \sigma'\theta_l)$ is not a zero of SL. Now let $\mathfrak{Z}$ be the prime differential ideal of $\mathfrak{F}_1\{y_1, \cdots, y_l\}$ with generic zero $(\sigma'\theta_1, \cdots, \sigma'\theta_l)$, and let $\mathfrak{Z}_1, \cdots, \mathfrak{Z}_s$ be the minimal prime differential ideal divisors of $\{\mathfrak{Z}, SL\}$ in $\mathfrak{F}_1\{y_1, \cdots, y_l\}$. If τ' is a specialization of σ' which can not be extended to a specialization of σ then $(\tau'\theta_1, \cdots, \tau'\theta_l)$ is a zero of $\mathfrak{Z}_j$ for some j. Let $(\psi_{j1}, \cdots, \psi_{jl})$ be a generic zero of this $\mathfrak{Z}_j$; then $(\psi_{j1}, \cdots, \psi_{jl})$ is a zero of $\mathfrak{Z}$. hence a specialization of $(\sigma'\theta_1, \cdots, \sigma'\theta_l)$ over $\mathfrak{F}_1$ and a fortiori over $\mathfrak{F}$, and therefore a specialization of $(\theta_1, \cdots, \theta_l)$ over $\mathfrak{F}$. But $(\psi_{j1}, \cdots, \psi_{jl})$ admits $(\tau'\theta_1, \cdots, \tau'\theta_l)$ as a specialization over $\mathfrak{F}_1$ and therefore over $\mathfrak{F}$, so that $(\theta_1, \cdots, \theta_l)$ is a specialization of $(\psi_{j1}, \cdots, \psi_{jl})$ over $\mathfrak{F}$. Thus $(\psi_{j1}, \cdots, \psi_{jl})$ is a generic specialization of $(\theta_1, \cdots, \theta_l)$ over $\mathfrak{F}$, so that there is a unique isomorphism σ'_j of $\mathfrak{F}_1$ over $\mathfrak{F}$ such that $\sigma'_j\theta_i = \psi_{ji}$ $(1 \leq i \leq l)$. Clearly τ' is a specialization of σ'_j, σ'_j is a specialization of σ', and because

$$\partial^0\sigma'_j\mathfrak{F}_1/\mathfrak{F}_1 = \operatorname{ord}\mathfrak{Z}_j < \operatorname{ord}\mathfrak{Z} = \partial^0\sigma'\mathfrak{F}_1/\mathfrak{F}_1$$

the latter specialization is nongeneric. This completes the proof of Corollary 3.

5. Algebraic sets. A subset $\mathfrak{M}^*$ of the group $\mathfrak{G}^*$ of all strong isomorphisms of $\mathfrak{H}$ over $\mathfrak{F}$ will be called an *irreducible set in* $\mathfrak{G}^*$ if $\mathfrak{M}^*$ contains an element σ^* such that $\mathfrak{M}^*$ is the set of all specializations of σ^*; any such σ^* will then be called a *generic element* of $\mathfrak{M}^*$, and the transcendence degree $\partial^0\mathfrak{H}\langle\sigma^*\mathfrak{H}\rangle/\mathfrak{H}$, which does not depend on the particular generic element σ^* employed, will be called the *dimension* of $\mathfrak{M}^*$ (notation: $\dim\mathfrak{M}^*$). It follows from Chapter I, Projosition 2, that if σ^* and τ^* are any two isomorphisms of $\mathfrak{H}$ over $\mathfrak{F}$ such that τ^* is a specialization of σ^* then $\partial^0\mathfrak{H}\langle\tau^*\mathfrak{H}\rangle/\mathfrak{H} \leq \partial^0\mathfrak{H}\langle\sigma^*\mathfrak{H}\rangle/\mathfrak{H}$, and we have equality here if and only if the specialization is generic. Consequently if $\mathfrak{M}^*$ and $\mathfrak{N}^*$ are irreducible sets in $\mathfrak{G}^*$ such that $\mathfrak{M}^* \supseteq \mathfrak{N}^*$ then $\dim\mathfrak{M}^* \geq \dim\mathfrak{N}^*$, and $\dim\mathfrak{M}^* = \dim\mathfrak{N}^*$ if and only if $\mathfrak{M}^* = \mathfrak{N}^*$. A subset $\mathfrak{M}^*$ of $\mathfrak{G}^*$ will be called an *algebraic set in* $\mathfrak{G}^*$ if $\mathfrak{M}^*$ is the union of a finite number of irreducible sets in $\mathfrak{G}^*$. It is easy to see that an algebraic set in $\mathfrak{G}^*$ can be written in one and only one way as the union of a finite set of irreducible sets in $\mathfrak{G}^*$ none of which contains another; these unique irreducible sets, which are the maximal irreducible sets contained in the algebraic set, will be called the *components* of the algebraic set.

It is a consequence of Corollary 2 of Proposition 9 that every nonempty algebraic set in $\mathfrak{G}^*$ has a nonempty intersection with the group $\mathfrak{G}$ of all automorphisms of $\mathfrak{H}$ over $\mathfrak{F}$, and that if $\mathfrak{M}^*$, $\mathfrak{N}^*$ are algebraic sets in $\mathfrak{G}^*$ with $\mathfrak{M}^* \neq \mathfrak{N}^*$ then $\mathfrak{M}^* \cap \mathfrak{G} \neq \mathfrak{N}^* \cap \mathfrak{G}$. Accordingly we shall call a subset $\mathfrak{M}$ of $\mathfrak{G}$ an *irreducible set in* $\mathfrak{G}$ or an *algebraic set in* $\mathfrak{G}$ if $\mathfrak{M}$ is the intersection with $\mathfrak{G}$ of, respectively, an irreducible set in $\mathfrak{G}^*$ or an algebraic set in $\mathfrak{G}^*$; this set in $\mathfrak{G}^*$, which is unique, we shall denote by $\mathfrak{M}^*$. If $\mathfrak{M}$ is an irreducible set in $\mathfrak{G}$ we define the *dimension* of $\mathfrak{M}$ (notation: $\dim \mathfrak{M}$) by the formula $\dim \mathfrak{M} = \dim \mathfrak{M}^*$, and define *generic element* of $\mathfrak{M}$ as a generic element of $\mathfrak{M}^*$ (so that a generic element of $\mathfrak{M}$ need not be an element of $\mathfrak{M}$); if σ^* is a generic element of $\mathfrak{M}$ we thus have $\dim \mathfrak{M} = \partial^0 \mathfrak{H}\langle \sigma^* \mathfrak{H}\rangle/\mathfrak{H}$. If $\mathfrak{M}$ and $\mathfrak{N}$ are two irreducible sets in $\mathfrak{G}$ with $\mathfrak{M} \supseteq \mathfrak{N}$ then $\dim \mathfrak{M} \geq \dim \mathfrak{N}$, and equality of dimension implies that $\mathfrak{M} = \mathfrak{N}$. An algebraic set in $\mathfrak{G}$ can be written in one and only one way as the union of a finite set of irreducible sets in $\mathfrak{G}$ none of which contains another; these unique irreducible sets, which are the maximal irreducible sets in $\mathfrak{G}$ contained in the algebraic set, will be called the *components* of the algebraic set. If $\mathfrak{M}$ is an algebraic set in $\mathfrak{G}$ and $\mathfrak{M}_1, \cdots, \mathfrak{M}_p$ are its components then $\mathfrak{M}^*_1, \cdots, \mathfrak{M}^*_p$ are the components of $\mathfrak{M}^*$.

PROPOSITION 10. *Every nonempty set of algebraic sets in* $\mathfrak{G}$ *has a minimal element.*

Proof. To each algebraic set $\mathfrak{M}$ we associate the sequence

$$k(\mathfrak{M}) = (k_d(\mathfrak{M}))_{0 \leq d < \infty},$$

where $k_d(\mathfrak{M})$ is the number of components of $\mathfrak{M}$ of dimension d; since the dimension of every irreducible set is $\leq \partial^0 \mathfrak{H}/\mathfrak{F}$ we have $k_d(\mathfrak{M}) = 0$ for all $d > \partial^0 \mathfrak{H}/\mathfrak{F}$. We introduce an order into the set of sequences $k(\mathfrak{M})$ by writing $k(\mathfrak{M}) \leq k(\mathfrak{N})$ whenever either $k(\mathfrak{M}) = k(\mathfrak{N})$ or $k(\mathfrak{M}) \neq k(\mathfrak{N})$ and the last nonzero difference $k_d(\mathfrak{N}) - k_d(\mathfrak{M})$ is positive. It is easy to see that if $\mathfrak{M} \subseteq \mathfrak{N}$ then $k(\mathfrak{M}) \leq k(\mathfrak{N})$; since it is obvious that in any nonempty set of algebraic sets in G there exists one for which the associated sequence is minimal, the proof is complete.

PROPOSITION 11. *If* $\mathfrak{M}_1, \cdots, \mathfrak{M}_p$ *are irreducible sets in* $\mathfrak{G}$ *then there exists an independent family of generic elements of* $\mathfrak{M}_1, \cdots, \mathfrak{M}_p$; *if* $\rho_i \in \mathfrak{M}^*_i$ $(1 \leq i \leq p)$ *then* $\rho_1, \cdots, \rho_p$ *are independent generic elements of* $\mathfrak{M}_1, \cdots, \mathfrak{M}_p$ *if and only if* $\partial^0 \mathfrak{H}\langle \rho_1 \mathfrak{H}, \cdots, \rho_p \mathfrak{H}\rangle/\mathfrak{H} = \sum_{i=1}^{p} \dim \mathfrak{M}_i$.

Proof. Let σ_i be a generic element of $\mathfrak{M}_i$, and let $\gamma_{i1}, \cdots, \gamma_{iq_i}$ be con-

stants such that $\mathfrak{k}_{\sigma_i} = \mathfrak{k}\langle\gamma_{i1}, \cdots, \gamma_{iq_i}\rangle$, $1 \leq i \leq p$. It is obvious that there exist generic specializations $(\delta_{i1}, \cdots, \delta_{iq_i})$ of $(\gamma_{i1}, \cdots, \gamma_{iq_i})$ over $\mathfrak{k}$, $1 \leq i \leq p$, such that

$$\partial^0 \mathfrak{k}\langle(\delta_{ij})_{1\leq i\leq p,\, 1\leq j\leq q_i}\rangle/\mathfrak{k} = \sum_{i=1}^{p} \partial^0 \mathfrak{k}\langle\delta_{i1}, \cdots, \delta_{iq_i}\rangle/\mathfrak{k}.$$

By Proposition 9 there exists a strong isomorphism τ_i of $\mathfrak{H}$ over $\mathfrak{F}$ such that τ_i is a specialization of σ_i and $\mathfrak{k}_{\tau_i} = \mathfrak{k}\langle\delta_{i1}, \cdots, \delta_{iq_i}\rangle$. Because

$$\begin{aligned}\partial^0 \mathfrak{H}\langle\tau_i\mathfrak{H}\rangle/\mathfrak{H} &= \partial^0 \mathfrak{k}\langle\delta_{i1}, \cdots, \delta_{iq_i}\rangle/\mathfrak{k} \\ &= \partial^0 \mathfrak{k}\langle\gamma_{i1}, \cdots, \gamma_{iq_i}\rangle/\mathfrak{k} = \partial^0 \mathfrak{H}\langle\sigma_i\mathfrak{H}\rangle/\mathfrak{H}\end{aligned}$$

it follows that

$$\partial^0 \mathfrak{H}\langle\tau_i\eta_1, \cdots, \tau_i\eta_n\rangle/\mathfrak{H} = \partial^0 \mathfrak{H}\langle\sigma_i\eta_1, \cdots, \sigma_i\eta_n\rangle/\mathfrak{H},$$

where $\eta_1, \cdots, \eta_n$ are elements of $\mathfrak{H}$ such that $\mathfrak{H} = \mathfrak{F}\langle\eta_1, \cdots, \eta_n\rangle$; therefore (Chapter I, Proposition 2) $(\tau_i\eta_1, \cdots, \tau_i\eta_n)$ is a generic specialization of $(\sigma_i\eta_1, \cdots, \sigma_i\eta_n)$ over $\mathfrak{H}$, so that τ_i is a generic specialization of σ_i and therefore a generic element of $\mathfrak{M}_i$. Because

$$\begin{aligned}\partial^0 \mathfrak{H}\langle\tau_1\mathfrak{H}, \cdots, \tau_p\mathfrak{H}\rangle/\mathfrak{H} &= \partial^0 \mathfrak{H}\langle(\delta_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}\rangle/\mathfrak{H} \\ &= \partial^0 \mathfrak{k}\langle(\delta_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}\rangle/\mathfrak{k} = \sum \partial^0 \mathfrak{k}\langle\delta_{i1}, \cdots, \delta_{iq_i}\rangle/\mathfrak{k} \\ &= \sum \partial^0 \mathfrak{k}_{\tau_i}/\mathfrak{k} = \sum \partial^0 \mathfrak{H}\langle\tau_i\mathfrak{H}\rangle/\mathfrak{H},\end{aligned}$$

we see from Proposition 7 that $\tau_1, \cdots, \tau_p$ are independent. If ρ_i is an element of $\mathfrak{M}^*_i$ $(1 \leq i \leq p)$ then $\rho_1, \cdots, \rho_p$ are independent (Proposition 7) if and only if $\partial^0 \mathfrak{H}\langle\rho_1\mathfrak{H}, \cdots, \rho_p\mathfrak{H}\rangle/\mathfrak{H} = \sum \partial^0 \mathfrak{H}\langle\rho_i\mathfrak{H}\rangle/\mathfrak{H}$, and are generic if and only if $\partial^0 \mathfrak{H}\langle\rho_i\mathfrak{H}\rangle/\mathfrak{H} = \dim \mathfrak{M}_i$ $(1 \leq i \leq p)$.

PROPOSITION 12. *If $\mathfrak{M}$ is an irreducible set in $\mathfrak{G}$ with generic element σ and if $\tau \in \mathfrak{G}$ then $\tau\mathfrak{M}$ and $\mathfrak{M}\tau$ and $\mathfrak{M}^{-1}$ are irreducible sets in $\mathfrak{G}$ with generic elements $\tau\sigma$ and $\sigma\tau$ and σ^{-1} respectively, and* $\dim \tau\mathfrak{M} = \dim \mathfrak{M}\tau = \dim \mathfrak{M}^{-1} = \dim \mathfrak{M}$.

Proof. The set consisting of the single element τ is obviously an irreducible set of dimension 0, and

$$\partial^0 \mathfrak{H}\langle\sigma\mathfrak{H}, \tau\mathfrak{H}\rangle/\mathfrak{H} = \partial^0 \mathfrak{H}\langle\sigma\mathfrak{H}\rangle/\mathfrak{H} = \dim \mathfrak{M} + 0;$$

therefore by Proposition 11 σ, τ are independent, so that (Proposition 8) every element of $\tau\mathfrak{M}$ is a specialization of $\tau\sigma$. Therefore if $\mathfrak{N}$ is the irreducible set in $\mathfrak{G}$ with generic element $\tau\sigma$ then $\tau\mathfrak{M} \subseteq \mathfrak{N}$. Similarly, $\tau^{-1}\mathfrak{N}$ is contained in the irreducible set in $\mathfrak{G}$ with generic element $\tau^{-1}(\tau\sigma) = \sigma$, that is in $\mathfrak{M}$, so that $\mathfrak{N} \subseteq \tau\mathfrak{M}$. Therefore $\tau\mathfrak{M} = \mathfrak{N}$, so that $\tau\mathfrak{M}$ is irreducible and

has generic element $\tau\sigma$. The proof for $\mathfrak{M}\tau$ is similar and for $\mathfrak{M}^{-1}$ is even simpler. Finally

$$\dim \tau\mathfrak{M} = \partial^0 \mathcal{G}\langle \tau\sigma\mathcal{G}\rangle/\mathcal{G} = \partial^0(\tau^{-1}\mathcal{G})\langle \sigma\mathcal{G}\rangle/\tau^{-1}\mathcal{G}$$
$$= \partial^0 \mathcal{G}\langle \sigma\mathcal{G}\rangle/\mathcal{G} = \dim \mathfrak{M},$$

$$\dim \mathfrak{M}\tau = \partial^0 \mathcal{G}\langle \sigma\tau\mathcal{G}\rangle/\mathcal{G} = \partial^0 \mathcal{G}\langle \sigma\mathcal{G}\rangle/\mathcal{G} = \dim \mathfrak{M},$$

and $\dim \mathfrak{M}^{-1} = \dim \mathfrak{M}$ by (2).

PROPOSITION 13. *If Φ is a nonempty set of algebraic sets in $\mathfrak{G}$ then $\bigcap_{\mathfrak{M}\in\Phi} \mathfrak{M}$ is an algebraic set in $\mathfrak{G}$.*

Proof. If Φ contains only one element the conclusion is obvious. Suppose next that Φ contains precisely two elements, say $\mathfrak{M}_1$ and $\mathfrak{M}_2$, and assume first that $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are irreducible; let σ_1 and σ_2 be generic elements of $\mathfrak{M}_1$ and $\mathfrak{M}_2$ respectively. Letting $\eta_1, \cdots, \eta_n$ be elements of $\mathcal{G}$ such that $\mathcal{G} = \mathcal{F}\langle \eta_1, \cdots, \eta_n\rangle$, we let $\mathbf{M}_i$ denote the prime differential ideal in $\mathcal{G}\{y_1, \cdots, y_n\}$ with generic zero $(\sigma_i\eta_1, \cdots, \sigma_i\eta_n)$, $i = 1, 2$. Let $\mathbf{N}_1, \cdots, \mathbf{N}_p$ be the minimal prime differential ideal divisors of the perfect differential ideal $\{\mathbf{M}_1, \mathbf{M}_2\}$ in $\mathcal{G}\{y_1, \cdots, y_n\}$, and let $(\zeta_{i1}, \cdots, \zeta_{in})$ be a generic zero of $\mathbf{N}_i$, $1 \leq i \leq p$. It is clear that $(\zeta_{i1}, \cdots, \zeta_{in})$ is a specialization of $(\sigma_1\eta_1, \cdots, \sigma_1\eta_n)$ and of $(\sigma_2\eta_1, \cdots, \sigma_2\eta_n)$ over $\mathcal{G}$ and a fortiori over $\mathcal{F}$, so that $(\zeta_{i1}, \cdots, \zeta_{in})$ is a specialization of $(\eta_1, \cdots, \eta_n)$ over $\mathcal{F}$. For every $\sigma \in \mathfrak{M}_1 \cap \mathfrak{M}_2$, $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a zero of $\{\mathbf{M}_1, \mathbf{M}_2\}$ and hence of some $\mathbf{N}_i$ and therefore is a specialization of $(\zeta_{i1}, \cdots, \zeta_{in})$ over $\mathcal{G}$ for some i. Let I be the set of all integers i with $1 \leq i \leq p$ such that $\mathfrak{M}_1 \cap \mathfrak{M}_2$ contains an element σ for which $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a specialization of $(\zeta_{i1}, \cdots, \zeta_{in})$. For each $i \in I$ there exists a $\sigma \in \mathfrak{M}_1 \cap \mathfrak{M}_2$ such that $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a specialization of $(\zeta_{i1}, \cdots, \zeta_{in})$ over $\mathcal{G}$ and a fortiori over $\mathcal{F}$, so that $(\eta_1, \cdots, \eta_n)$ is a specialization of $(\zeta_{i1}, \cdots, \zeta_{in})$ over $\mathcal{F}$. By the above, $(\zeta_{i1}, \cdots, \zeta_{in})$ is then a generic specialization of $(\eta_1, \cdots, \eta_n)$ over $\mathcal{F}$, so that there exists a unique isomorphism τ_i of $\mathcal{G}$ over $\mathcal{F}$ such that $\tau_i\eta_1 = \zeta_{i1}, \cdots, \tau_i\eta_n = \zeta_{in}$. Let $\mathfrak{N}_i$ be the irreducible set in $\mathfrak{G}$ with generic element τ_i. Each element of $\mathfrak{M}_1 \cap \mathfrak{M}_2$ is a specialization of some τ_i and therefore belongs to $\bigcup_{i\in I} \mathfrak{N}_i$. Conversely, each element of $\bigcup_{i\in I} \mathfrak{N}_i$ is a specialization of some τ_i and hence also of σ_1 and σ_2 and therefore belongs to $\mathfrak{M}_1 \cap \mathfrak{M}_2$. Thus $\mathfrak{M}_1 \cap \mathfrak{M}_2 = \bigcup_{i\in I} \mathfrak{N}_i$, so that $\mathfrak{M}_1 \cap \mathfrak{M}_2$ is algebraic.

Continuing with the case in which Φ consists of $\mathfrak{M}_1$ and $\mathfrak{M}_2$, now abandon the assumption that $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are irreducible. Denoting the components

of $\mathfrak{M}_i$ by $\mathfrak{M}_{i1}, \cdots, \mathfrak{M}_{ip_i}$ $(i = 1, 2)$, we have $\mathfrak{M}_1 \cap \mathfrak{M}_2 = \bigcup_{j,k} \mathfrak{M}_{1j} \cap \mathfrak{M}_{2k}$. By what has already been proved, each $\mathfrak{M}_{1j} \cap \mathfrak{M}_{2k}$ is algebraic, so that $\mathfrak{M}_1 \cap \mathfrak{M}_2$ is too. The proposition thus holds when Φ consists of two algebraic sets. The extension to the case in which Φ consists of any finite number of algebraic sets is immediate by induction.

Now let Φ be arbitrary. Let Ψ be the set of all intersections of finite nonempty subsets of Φ. By the case already known, each element of Ψ is an algebraic set in $\mathfrak{G}$. By Proposition 10, Ψ has a minimal element, say $\mathfrak{N}$. It is obvious that $\mathfrak{M} \cap \mathfrak{N} \in \Psi$ for each $\mathfrak{M} \in \Phi$; since $\mathfrak{M} \cap \mathfrak{N} \subseteq \mathfrak{N}$ and $\mathfrak{N}$ is minimal in Ψ this means that $\mathfrak{M} \cap \mathfrak{N} = \mathfrak{N}$, that is $\mathfrak{N} \subseteq \mathfrak{M}$. Therefore $\mathfrak{N} = \bigcap_{\mathfrak{M} \in \Phi} \mathfrak{M}$, and the latter is algebraic.

6. Algebraic groups. By an *algebraic group* in $\mathfrak{G}$ we shall mean a subset of $\mathfrak{G}$ which is a subgroup of $\mathfrak{G}$ and at the same time an algebraic set in $\mathfrak{G}$. If $\mathfrak{H}$ is an algebraic group in $\mathfrak{G}$ and $\mathfrak{H}^*$ is the algebraic set in $\mathfrak{G}^*$ such that $\mathfrak{H} = \mathfrak{G} \cap \mathfrak{H}^*$ then $\mathfrak{H}^*$ is a subgroup of $\mathfrak{G}^*$. To prove this we observe that for two independent generic points σ^*, τ^* of two not necessarily distinct components of $\mathfrak{H}^*$ (see Proposition 11) we have $\sigma^{*-1}\tau^* \in \mathfrak{H}^*$, for otherwise by Proposition 9 we could find elements $\sigma, \tau \in \mathfrak{H}$ such that $\sigma^{-1}\tau \notin \mathfrak{H}$; it follows from Proposition 6 that $\sigma^{-1}\tau \in \mathfrak{H}^*$ whenever $\sigma, \tau \in \mathfrak{H}^*$, so that $\mathfrak{H}^*$ is a group.

THEOREM 1. *Let $\mathfrak{H}$ be an algebraic group in $\mathfrak{G}$. The components of $\mathfrak{H}$ are pairwise disjoint; the component $\mathfrak{H}^0$ of $\mathfrak{H}$ which contains the identity automorphism ι is a normal algebraic subgroup of $\mathfrak{H}$ of finite index; the components of $\mathfrak{H}$ are the cosets of $\mathfrak{H}^0$ in $\mathfrak{H}$.*

Proof. Let $\mathfrak{H}_1$, $\mathfrak{H}_2$ be components of $\mathfrak{H}$ which contain ι and let σ^*_1, σ^*_2 be independent generic elements of $\mathfrak{H}_1$, $\mathfrak{H}_2$. Since $\sigma^*_1\sigma^*_2$ belongs to $\mathfrak{H}^*$ it belongs to some component $\mathfrak{H}^*_0$ of $\mathfrak{H}^*$. By Proposition 8, $\sigma^*_1 = \sigma^*_1\iota$ is a specialization of $\sigma^*_1\sigma^*_2$ so that $\mathfrak{H}^*_1 \subseteq \mathfrak{H}^*_0$ and therefore $\mathfrak{H}^*_1 = \mathfrak{H}^*_0$, and similarly $\sigma^*_2 = \iota\sigma^*_2$ is a specialization of $\sigma^*_1\sigma^*_2$ so that $\mathfrak{H}^*_2 \subseteq \mathfrak{H}^*_0$ and $\mathfrak{H}^*_2 = \mathfrak{H}^*_0$; therefore $\mathfrak{H}^*_1 = \mathfrak{H}^*_2$ so that $\mathfrak{H}_1 = \mathfrak{H}_2$. Thus precisely one of the components of $\mathfrak{H}$ contains ι; we denote this component by $\mathfrak{H}^0$. Now let $\mathfrak{H}'$ be any component of $\mathfrak{H}$; if σ' is any element of $\mathfrak{H}'$ then (Proposition 12) $\sigma'^{-1}\mathfrak{H}'$ is an irreducible subset of $\mathfrak{H}$ containing ι so that $\sigma'^{-1}\mathfrak{H}' \subseteq \mathfrak{H}^0$ and $\mathfrak{H}' \subseteq \sigma'\mathfrak{H}^0$. Since (Proposition 12) $\sigma'\mathfrak{H}^0$ is an irreducible subset of $\mathfrak{H}$ and $\mathfrak{H}'$ is a component of $\mathfrak{H}$ this implies that $\mathfrak{H}' = \sigma'\mathfrak{H}^0$. Similarly we find that $\mathfrak{H}' = \mathfrak{H}^0\sigma'$, so that $\sigma'\mathfrak{H}^0 = \mathfrak{H}^0\sigma'$. If we choose $\mathfrak{H}'$ as $\mathfrak{H}^0$ we see that $\sigma\mathfrak{H}^0 = \mathfrak{H}^0$

for every $\sigma \in \mathfrak{H}^0$ so that $\mathfrak{H}^0$ is a subgroup of $\mathfrak{H}$; since $\sigma'\mathfrak{H}^0 = \mathfrak{H}^0\sigma'$ for every $\sigma' \in \mathfrak{H}$, $\mathfrak{H}^0$ is a normal subgroup of $\mathfrak{H}$. By the above, the components of $\mathfrak{H}$ are the cosets of $\mathfrak{H}^0$ in $\mathfrak{H}$, and are therefore pairwise disjoint.

COROLLARY 1. *The components of an algebraic group in $\mathfrak{G}$ all have the same dimension.*

Proof. This follows from the Theorem and Proposition 12.

By the *dimension* of an algebraic group $\mathfrak{H}$ in $\mathfrak{G}$ we shall mean the dimension of any one of its components. The component containing ι we shall call the *component of the identity* of $\mathfrak{H}$, and we shall always denote it by $\mathfrak{H}^0$.

COROLLARY 2. *An algebraic group in $\mathfrak{G}$ is irreducible if and only if it has no algebraic subgroup of finite index > 1.*

Proof. If $\mathfrak{H}$ is not irreducible then $\mathfrak{H}^0$ is an algebraic subgroup of $\mathfrak{H}$ of finite index > 1. Conversely, if $\mathfrak{K}$ is an algebraic subgroup of $\mathfrak{H}$ of finite index > 1 then so is the component of the identity $\mathfrak{K}^0$ of $\mathfrak{K}$; as $\mathfrak{H}$ is the union of the left cosets of $\mathfrak{K}^0$ in $\mathfrak{H}$ and these are irreducible sets (Proposition 12) none of which contains another, $\mathfrak{H}$ is not irreducible.

PROPOSITION 14. *Let $\mathfrak{h}$ be a subgroup of $\mathfrak{G}$ which is contained in at least one algebraic set in $\mathfrak{G}$. Then the intersection $\mathfrak{H}$ of all the algebraic sets in $\mathfrak{G}$ which contain $\mathfrak{h}$ is an algebraic group in $\mathfrak{G}$.*

Proof. By Proposition 13, $\mathfrak{H}$ is an algebraic set in $\mathfrak{G}$. Obviously $\mathfrak{H}$ can be characterized as the smallest algebraic set in $\mathfrak{G}$ which contains $\mathfrak{h}$. If σ is any element of $\mathfrak{h}$ then $\mathfrak{H}\sigma$, which by Proposition 12 is an algebraic set, is obviously the smallest algebraic set containing $\mathfrak{h}\sigma$. Since $\mathfrak{h}\sigma = \mathfrak{h}$ this means that $\mathfrak{H}\sigma = \mathfrak{H}$, so that $\mathfrak{H}\mathfrak{h} = \mathfrak{H}$. Now let σ be any element of $\mathfrak{H}$. Obviously $\sigma\mathfrak{H}$ is the smallest algebraic set containing $\sigma\mathfrak{h}$; but $\sigma\mathfrak{h} \subseteq \mathfrak{H}\mathfrak{h} = \mathfrak{H}$, so that $\sigma\mathfrak{H} \subseteq \mathfrak{H}$. By Proposition 12, $\mathfrak{H}^{-1}$ is algebraic and therefore is the smallest algebraic set containing $\mathfrak{h}^{-1}$; since $\mathfrak{h}^{-1} = \mathfrak{h}$ this means that $\mathfrak{H}^{-1} = \mathfrak{H}$, so that $\mathfrak{H}$ is a group.

PROPOSITION 15. *Let $\mathfrak{H}$ be an algebraic group in $\mathfrak{G}$ and let $\mathfrak{h}$ be a subgroup of $\mathfrak{H}$ such that $\mathfrak{H} - \mathfrak{h}$ is contained in the union of a finite family of irreducible sets in $\mathfrak{G}$ each of dimension $<$ dim $\mathfrak{H}$. Then $\mathfrak{h} = \mathfrak{H}$.*

Proof. Expressing $\mathfrak{H}$ as the union of cosets of $\mathfrak{h}$, $\mathfrak{H} = \bigcup_{i \in I} \sigma_i\mathfrak{h}$ $(\sigma_{i_0} = \iota)$, we see that if there were more than one coset there would exist an $i_1 \in I$

with $i_0 \neq i_1$; for this i_1 we would have $\sigma_{i_1}\mathfrak{h} \subseteq \mathfrak{H} - \mathfrak{h}$, so that $\sigma_{i_1}\mathfrak{h}$ would be contained in a finite union of irreducible sets each of dimension $< \dim \mathfrak{H}$, whence (Proposition 12) so would $\mathfrak{h}$. This would imply that the same is true of $\mathfrak{H} = \mathfrak{h} \cup (\mathfrak{H} - \mathfrak{h})$, which is impossible.

PROPOSITION 16. *Let $\mathfrak{H}$ be an algebraic group in $\mathfrak{G}$, $\mathfrak{K}$ an algebraic subgroup of $\mathfrak{H}$, $\mathfrak{N}$ a normal algebraic subgroup of $\mathfrak{H}$. Then $\mathfrak{K}\mathfrak{N}$ is an algebraic group in $\mathfrak{G}$.*

Proof. $\mathfrak{K}\mathfrak{N}$ is a group; we must prove it is algebraic. Let (Proposition 11) $\kappa_1, \cdots, \kappa_r, \nu_1, \cdots, \nu_s$ be independent generic elements of the r components of $\mathfrak{K}$ and the s components of $\mathfrak{N}$, let $\mathfrak{M}_{ij}$ be the irreducible set in $\mathfrak{G}$ with generic element $\kappa_i\nu_j$, and let $\mathfrak{M}$ be the intersection of all the algebraic sets in $\mathfrak{G}$ which contain $\mathfrak{K}\mathfrak{N}$. By Proposition 8, $\mathfrak{K}\mathfrak{N} \subseteq \bigcup \mathfrak{M}_{ij}$, so that (Proposition 14) $\mathfrak{M}$ is an algebraic group and $\mathfrak{M} \subseteq \bigcup \mathfrak{M}_{ij}$. If we had $\kappa_i\nu_j \notin \mathfrak{M}^*$ then (Proposition 9) there would exist $\kappa' \in \mathfrak{K}$, $\nu'_j \in \mathfrak{N}$ such that $\kappa'_i\nu'_j \notin \mathfrak{M}$, which is impossible; it follows that $\mathfrak{M} = \bigcup \mathfrak{M}_{ij}$, so that $\dim \mathfrak{M} \geq \dim \mathfrak{M}_{ij}$. By Corollary 1 to Proposition 9, the set of elements of $\mathfrak{M}_{ij}$ which do not belong to $\mathfrak{K}\mathfrak{N}$ lies in an algebraic set properly contained in $\mathfrak{M}_{ij}$, that is, in a finite union of irreducible sets all of dimension $< \dim \mathfrak{M}_{ij}$. It follows that $\mathfrak{M} - \mathfrak{K}\mathfrak{N}$ is contained in a finite union of irreducible sets each of dimension $< \dim \mathfrak{M}$, so that (Proposition 15) $\mathfrak{K}\mathfrak{N} = \mathfrak{M}$, whence $\mathfrak{K}\mathfrak{N}$ is algebraic.

PROPOSITION 17. *Let $\mathfrak{K}$ be an algebraic subgroup of an algebraic group $\mathfrak{H}$ in $\mathfrak{G}$, and let $N(\mathfrak{K})$ be the normalisor of $\mathfrak{K}$ in $\mathfrak{H}$. Then $N(\mathfrak{K})$ is an algebraic group in $\mathfrak{G}$.*

Proof. The intersection $\mathfrak{N}$ of all the algebraic sets containing $N(\mathfrak{K})$ is an algebraic group (Proposition 14); we shall show that $N(\mathfrak{K}) = \mathfrak{N}$, thereby proving that $N(\mathfrak{K})$ is algebraic. Let $\mathfrak{N}_0$ be any component of $\mathfrak{N}$ and let σ_0 be a generic element of $\mathfrak{N}_0$. Suppose $\sigma_0\mathfrak{K}^*\sigma_0^{-1} \subseteq \mathfrak{K}^*$. Using Proposition 9 we see that there exists an element $\tau \in \mathfrak{K}$ such that $\sigma_0\tau\sigma_0^{-1} \notin \mathfrak{K}^*$. Letting $\eta_1, \cdots, \eta_n$ be elements of $\mathfrak{G}$ such that $\mathfrak{G} = \mathfrak{F}\langle\eta_1, \cdots, \eta_n\rangle$, we see that therefore there exists a $Q \in \mathfrak{G}\{w_1, \cdots, w_n\}$ which vanishes at $(\sigma\eta_1, \cdots, \sigma\eta_n)$ for every $\sigma \in \mathfrak{K}$ but which does not vanish at $(\sigma_0\tau\sigma_0^{-1}\eta_1, \cdots, \sigma_0\tau\sigma_0^{-1}\eta_n)$. By Corollary 1 to Proposition 9 there exists a $P \in \mathfrak{G}\{y_1, \cdots, y_n\}$ which does not vanish at $(\sigma_0\eta_1, \cdots, \sigma_0\eta_n)$ such that whenever σ'_0 is a specialization of σ_0 for which $P(\sigma'_0\eta_1, \cdots, \sigma'_0\eta_n) \neq 0$ then $Q(\sigma'_0\tau\sigma'^{-1}_0\eta_1, \cdots, \sigma'_0\tau\sigma'^{-1}_0\eta_n) \neq 0$, that is then $\sigma'_0\tau\sigma'^{-1}_0 \notin \mathfrak{K}$, so that $\sigma'_0 \notin N(\mathfrak{K})$. Thus $P(\sigma\eta_1, \cdots, \sigma\eta_n) = 0$ for every $\sigma \in N(\mathfrak{K}) \cap \mathfrak{N}_0$. It easily follows that $N(\mathfrak{K}) \cap \mathfrak{N}_0$ is contained in an algebraic

set properly contained in $\mathfrak{N}_0$, so that $N(\mathfrak{R})$ is contained in an algebraic set properly contained in $\mathfrak{N}$. This contradicts the definition of $\mathfrak{N}$, and proves that $\sigma_0 \mathfrak{R}^* \sigma_0^{-1} \subseteq \mathfrak{R}^*$. From this it easily follows, using Corollary 1 to Proposition 9, that $\sigma \mathfrak{R} \sigma^{-1} \subseteq \mathfrak{R}$ for all $\sigma \in \mathfrak{N}$ save possibly those in a finite union of irreducible sets all of dimension $<$ dim $\mathfrak{N}$, that is $\mathfrak{N} - \mathfrak{N}(K)$ is contained in such a union. From Proposition 15 it now follows that $N(\mathfrak{R}) = \mathfrak{N}$.

Chapter III. Galois theory of strongly normal extensions.

We recall that the conditions and conventions set forth at the beginning of Chapter II remain in force in the present chapter.

1. Normal extensions. We recall (Kolchin [3], Chapter III) two definitions: 1) a set of isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ is said to be *abundant* if for every differential field $\mathcal{F}_1$ between $\mathcal{F}$ and $\mathcal{G}$ and every element α in $\mathcal{G}$ but not in $\mathcal{F}_1$ there exists an isomorphism, in the set, which leaves invariant each element of $\mathcal{F}_1$ but which does not leave α invariant; 2) $\mathcal{G}$ is said to be a *normal* extension of $\mathcal{F}$ if the set of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ is abundant. We now introduce the following definition: $\mathcal{G}$ is said to be a *weakly normal* extension of $\mathcal{F}$ if for every element α in $\mathcal{G}$ but not in $\mathcal{F}$ there exists an automorphism of $\mathcal{G}$ over $\mathcal{F}$ which does not leave α invariant. It is clear that if $\mathcal{G}$ is normal over $\mathcal{F}$ then $\mathcal{G}$ is weakly normal over $\mathcal{F}$, and that $\mathcal{G}$ is normal over $\mathcal{F}$ if and only if $\mathcal{G}$ is weakly normal over every differential field between $\mathcal{F}$ and $\mathcal{G}$. Whether $\mathcal{G}$ can be weakly normal over $\mathcal{F}$ without being normal over $\mathcal{F}$ is an open question.[7]

The following result was proved in Kolchin [3], § 16.

THEOREM 1. *Let $\mathcal{G}$ be a normal extension of $\mathcal{F}$, let $\mathfrak{G}$ be an abundant group of automorphisms of $\mathcal{G}$ over $\mathcal{F}$* (not necessarily the group of all such automorphisms), *and for each differential field $\mathcal{F}_1$ between $\mathcal{F}$ and $\mathcal{G}$ let*

[7] If we relax the conditions on $\mathcal{G}$ and $\mathcal{F}$ by dropping the requirement that every constant in $\mathcal{G}$ belong to $\mathcal{F}$ then the answer to this question is affirmative, as is shown by the following example. Let $\mathcal{F}$ be the field of all algebraic numbers, let θ be a transcendental number, let $\mathcal{G} = \mathcal{F}(\theta)$, and make $\mathcal{F}$ and $\mathcal{G}$ into ordinary differential fields by defining $\delta_1 a = 0$ for every $a \in \mathcal{G}$. The automorphisms of $\mathcal{G}$ over $\mathcal{F}$ may be identified with the fractional linear substitutions $\theta \to (a\theta + b)(c\theta + d)^{-1}$ with $a, b, c, d \in \mathcal{F}$ and $ad - bc \neq 0$. The only elements of $\mathcal{G}$ invariant under $\theta \to \theta + 1$ are those of $\mathcal{F}$, so that $\mathcal{G}$ is weakly normal over $\mathcal{F}$ (in the relaxed sense). But $\theta \notin \mathcal{F}(\theta^3 + \theta)$, and the only fractional linear substitution leaving $\theta^3 + \theta$ invariant is $\theta \to \theta$, which leaves θ invariant, too: therefore $\mathcal{G}$ is not normal over $\mathcal{F}$ (in the relaxed sense).

$\mathfrak{G}(\mathfrak{F}_1)$ denote the group of all elements of $\mathfrak{G}$ which leave invariant each element of $\mathfrak{F}_1$. Then for each $\mathfrak{F}_1$ the field of all elements of $\mathfrak{H}$ invariant under every element of $\mathfrak{G}(\mathfrak{F}_1)$ is $\mathfrak{F}_1$, so that $\mathfrak{F}_1 \to \mathfrak{G}(\mathfrak{F}_1)$ is a one-to-one mapping of the set of all differential fields between $\mathfrak{F}$ and $\mathfrak{H}$ onto a certain set of subgroups of $\mathfrak{G}$. A necessary and sufficient condition that $\mathfrak{G}(\mathfrak{F}_1)$ be a normal subgroup of $\mathfrak{H}$ is that $\sigma\mathfrak{F}_1 \subseteq \mathfrak{F}_1$ for every $\sigma \in \mathfrak{G}$, and when this condition is satisfied then the mapping which to each element of $\mathfrak{G}$ assigns its restriction to $\mathfrak{F}_1$ is a homomorphism with kernel $\mathfrak{G}(\mathfrak{F}_1)$ of $\mathfrak{G}$ onto an abundant group of automorphisms of $\mathfrak{F}_1$ over $\mathfrak{F}$.

We shall give an example which will show that even when $\mathfrak{G}$ is taken as the group of *all* automorphisms of $\mathfrak{H}$ over $\mathfrak{F}$ and $\mathfrak{G}(\mathfrak{F}_1)$ is a normal subgroup of $\mathfrak{G}$ (so that $\mathfrak{F}_1$ is a normal extension of $\mathfrak{F}$) the factor group $\mathfrak{G}/\mathfrak{G}(\mathfrak{F}_1)$ need not be isomorphic to the group of all automorphisms of $\mathfrak{F}_1$ over $\mathfrak{F}$, but merely to an abundant subgroup thereof; the example will also show that it is possible for an intermediate differential field $\mathfrak{F}_1$ that $\mathfrak{F}_1$ be a normal extension of $\mathfrak{F}$ and $\mathfrak{G}(\mathfrak{F}_1)$ fail to be a normal subgroup of $\mathfrak{G}$. To discuss this example we shall need the following lemma.

LEMMA 1. *Let p, q be integers not both zero, let a, b, c, d be nonzero elements of $\mathfrak{C}$, let X_1, X_2, X_3, X_4 be indeterminates, and let $f \in \mathfrak{C}(X_1, X_2, X_3, X_4)$. If f is invariant under the substitution*

$$(1) \qquad (X_1, X_2, X_3, X_4) \to (X_1 + d, aX_2, bX_3, cX_2^p X_3^q X_4)$$

then $f \in \mathfrak{C}(X_2, X_3)$.

Proof. We suppose as we may that $f \neq 0$; then f is uniquely expressible in the form

$$f = \sum_{i=0}^{m} g_i X_4^i / \sum_{j=0}^{m} h_j X_4^j \quad (g_i, h_j \in \mathfrak{C}(X_1, X_2, X_3),\ g_m \neq 0,\ h_n = 1),$$

where $\sum g_i X_4^i$ and $\sum h_j X_4^j$ have no common factor as polynomials in X_4. Because of the invariance of f under the indicated substitution we find that

$$g_i(X_1 + d, aX_2, bX_3) = g_i(X_1, X_2, X_3)\,(cX_2^p X_3^q)^{n-i},$$

$$h_j(X_1 + d, aX_2, bX_3) = h_j(X_1, X_2, X_3)\,(cX_2^p X_3^q)^{n-j}.$$

But the degree in X_2 of the numerator of any nonzero element $\phi \in \mathfrak{C}(X_1, X_2, X_3)$ minus the degree in X_2 of the denominator of ϕ is obviously invariant under the substitution

$$(2) \qquad (X_1, X_2, X_3) \to (X_1 + d, aX_2, bX_3);$$

therefore $p(n-i)=0$ for every i such that $g_i \neq 0$. Similarly, regarding degrees in X_3 instead of X_2, we see that $q(n-i)=0$ whenever $g_i \neq 0$. Since p and q are not both 0 this implies that $g_i = 0$ whenever $i \neq n$. In the same way we find that $h_j = 0$ whenever $j \neq n$. Therefore $m = n = 0$, and $f \in \mathfrak{C}(X_1, X_2, X_3)$; thus we may write $f = f(X_1, X_2, X_3)$, and f is invariant under the substitution (2).

Now the set $\mathfrak{H}$ of all matrices

$$\tau(\alpha, \beta, \delta) = \begin{pmatrix} 1 & \delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & \beta \end{pmatrix} \quad (\alpha, \beta, \delta \in \mathfrak{C}, \alpha\beta \neq 0)$$

such that $f(X_1 + \delta, \alpha X_2, \beta X_3) = f(X_1, X_2, X_3)$, is an algebraic matic group, and $\tau(a, b, d) \in \mathfrak{H}$. It follows (Kolchin [5], § 3, Lemma 1) that $\tau(1, 1, d) \in \mathfrak{H}$, that is, $f(X_1 + d, X_2, X_3) = f(X_1, X_2, X_3)$. Since $d \neq 0$ it is easy to conclude that f is free of X_1, that is $f \in \mathfrak{C}(X_2, X_3)$.

EXAMPLE. Let $\mathfrak{F}$ be the field of all complex numbers and let $\mathfrak{G}$ be the field $\mathfrak{F}(x, e^x, e^{ix}, e^{x^2})$ obtained by the field adjunction to $\mathfrak{F}$ of the functions x, e^x, e^{ix}, e^{x^2} of the complex variable x $(i = \sqrt{-1})$; with respect to the operator d/dx, $\mathfrak{G}$ is an ordinary differential field with field of constants $\mathfrak{F}$.

We shall first show that if $\mathfrak{F}_1$ is a differential field between $\mathfrak{F}$ and $\mathfrak{G}$ then either $\mathfrak{F}_1$ is between $\mathfrak{F}\langle x\rangle$ and $\mathfrak{G}$ or else $\mathfrak{F}_1$ is between $\mathfrak{F}$ and $\mathfrak{F}\langle e^x, e^{ix}\rangle$. Indeed, suppose $x \notin \mathfrak{F}_1$, and let $\theta \in \mathfrak{F}_1$; we must prove that $\theta \in \mathfrak{F}\langle e^x, e^{ix}\rangle$. Since $x \notin \mathfrak{F}_1$ there exists an isomorphism of $\mathfrak{G}$ over $\mathfrak{F}_1$ under which x is not invariant; it easily follows that there exist nonzero complex numbers a_0, b_0, c_0, d such that $(x + d, a_0 e^x, b_0 e^{ix}, c_0 e^{(x+d)^2})$ is a generic specialization of $(x, e^x, e^{ix}, e^{x^2})$ over $\mathfrak{F}_1$. Letting $c = c_0 e^{d^2}$ and letting f be an element of $\mathfrak{F}(X_1, X_2, X_3, X_4)$ such that $\theta = f(x, e^x, e^{ix}, e^{x^2})$, we see that

$$f(x + d, a_0 e^x, b_0 e^{ix}, c e^{2dx} e^{x^2}) = f(x, e^x, e^{ix}, e^{x^2}). \tag{3}$$

If f were not free of X_4 this would mean that $e^x, e^{ix}, e^{2dx}, e^{x^2}$ are algebraically dependent over $\mathfrak{F}(x)$, and this would imply that $x, ix, 2dx, x^2$ are linearly dependent over the ring of integers, that is, there would exist integers p, q, r with $r \neq 0$ and p, q not both 0 such that $2d = (p + iq)/r$. Choosing complex numbers a, b such that $a_0 = a^r$, $b_0 = b^r$ we would then have

$$f(x + d, a^r e^x, b^r e^{ix}, c e^{[(p+iq)/r]x} e^{x^2}) = f(x, e^x, e^{ix}, e^{x^2}).$$

Since $e^{x/r}, e^{ix/r}, e^{x^2}$ are algebraically independent over $\mathfrak{F}(x)$, this would imply

that $f(X_1, X_2^r, X_3^r, X_4)$ is invariant under the substitution (1), so that by Lemma 1 we would have $f \in \mathfrak{F}(X_2, X_3)$. It follows that f is free of X_4. Therefore we may write (3) in the form

$$f(x + d, a_0 e^x, b_0 e^{ix}, c_0 e^x e^{x^2}) = f(x, e^x, e^{ix}, e^{x^2}),$$

whence f is invariant under the substitution (1) (with $(a_0, b_0, c_0, d, 1, 0)$ instead of (a, b, c, d, p, q)). It follows from Lemma 1 that $f \in \mathfrak{F}(X_2, X_3)$, so that $\theta \in \mathfrak{F}\langle e^x, e^{ix}\rangle$. This completes the proof that $\mathfrak{F}_1$ is either between $\mathfrak{F}\langle x\rangle$ and $\mathfrak{G}$ or between $\mathfrak{F}$ and $\mathfrak{F}\langle e^x, e^{ix}\rangle$.

It is easy to see that for each choice of nonzero complex numbers a, b, c and integers p, q there exists a unique automorphism $\sigma = \sigma(p, q, a, b, c)$ of $\mathfrak{G}$ over $\mathfrak{F}$ such that $\sigma x = x + \frac{1}{2}(p + iq)$, $\sigma e^x = ae^x$, $\sigma e^{ix} = be^{iz}$, $\sigma e^{x^2} = ce^{px} e^{qix} e^{x^2}$, and every automorphism of $\mathfrak{G}$ over $\mathfrak{F}$ is of this form; let us denote the group of all these automorphisms by $\mathfrak{G}$.

Let $\mathfrak{F}_1$ be between $\mathfrak{F}$ and $\mathfrak{G}$. If $\mathfrak{F}_1$ is between $\mathfrak{F}\langle x\rangle$ and $\mathfrak{G}$ then, since $\mathfrak{G} = \mathfrak{F}\langle x\rangle\langle e^x, e^{ix}, e^{x^2}\rangle$ is obviously a Picard-Vessiot extension of $\mathfrak{F}\langle x\rangle$, $\mathfrak{G}$ is normal over $\mathfrak{F}\langle x\rangle$ and therefore weakly normal over $\mathfrak{F}_1$. Suppose, then, that $\mathfrak{F}_1$ is between $\mathfrak{F}$ and $\mathfrak{F}\langle e^x, e^{ix}\rangle$, and let α be an element of $\mathfrak{G}$ but not of $\mathfrak{F}_1$. If $\alpha \notin \mathfrak{F}_1\langle e^x, e^{ix}\rangle$ then, by Lemma 1 and the algebraic independence of e^x, e^{ix}, e^{x^2} over $\mathfrak{F}(x)$, the automorphism $\sigma(1, 0, 1, 1, 1)$ does not leave α invariant but obviously leaves each element of $\mathfrak{F}_1$ invariant; suppose, then, that $\alpha \in \mathfrak{F}\langle e^x, e^{ix}\rangle$. Since $\mathfrak{F}\langle e^x, e^{ix}\rangle$ is obviously a Picard-Vessiot extension of $\mathfrak{F}$, there exists an automorphism σ_0 of $\mathfrak{F}\langle e^x, e^{ix}\rangle$ over $\mathfrak{F}_1$ such that $\sigma_0\alpha \neq \alpha$; now there exist nonzero complex numbers a, b such that $\sigma_0 e^x = ae^x$, $\sigma_0 e^{ix} = be^{ix}$. It is clear that the element $\sigma(0, 0, a, b, 1)$ of $\mathfrak{G}$ is an extension of σ_0, and therefore does not leave α invariant but leaves each element of $\mathfrak{F}_1$ invariant. Thus again $\mathfrak{G}$ is a weakly normal extension of $\mathfrak{F}_1$. Since $\mathfrak{F}_1$ is arbitrary between $\mathfrak{F}$ and $\mathfrak{G}$, $\mathfrak{G}$ is a normal extension of $\mathfrak{F}$.

Now let $\mathfrak{F}_2 = \mathfrak{F}\langle x, e^x, e^{x^2}\rangle$, $\mathfrak{F}_3 = \mathfrak{F}\langle x\rangle$. It is clear that $\sigma\mathfrak{F}_3 = \mathfrak{F}_3$ for every $\sigma = \sigma(p, q, a, b, c) \in \mathfrak{G}$, so that $\mathfrak{G}(\mathfrak{F}_3)$ is a normal subgroup of $\mathfrak{G}$; the restrictions to $\mathfrak{F}_3$ of the elements of $\mathfrak{G}$ are the automorphisms $\sigma(d)$ of $\mathfrak{F}_3$ over $\mathfrak{F}$ defined by

$$\sigma(d)x = x + d,$$

with d an arbitrary number of the form $\frac{1}{2}(p + iq)$, p and q being integers, and the group of all these restrictions is not the group of all automorphisms of $\mathfrak{F}_3$ over $\mathfrak{F}$, that is, is not the group of all automorphisms $\sigma(d)$ with d an arbitrary complex number, but is merely an abundant subgroup thereof.

Finally, there exists an automorphism $\sigma \in \mathfrak{G}$ such that $\sigma F_2 \subsetneq F_2$ (for

example $\sigma = \sigma(0, 1, 1, 1, 1)$), so that $\mathfrak{G}(\mathcal{F}_2)$ is not a normal subgroup of $\mathfrak{G}$; nevertheless $\mathcal{F}_2$ is a normal extension of $\mathcal{F}$, as can be shown by the method used to prove that $\mathcal{G}$ is a normal extension of $\mathcal{F}$.

2. Strongly normal extensions. We make the following definition: $\mathcal{G}$ is said to be a *strongly normal* extension of $\mathcal{F}$ if every isomorphism of $\mathcal{G}$ over $\mathcal{F}$ is strong. It is obvious that if $\mathcal{G}$ is strongly normal over $\mathcal{F}$ then $\mathcal{G}$ is strongly normal over every differential field between $\mathcal{F}$ and $\mathcal{G}$.

PROPOSITION 1. *A necessary and sufficient condition that $\mathcal{G}$ be a strongly normal extension of $\mathcal{F}$ is that $\chi\mathcal{G} \subseteq \mathcal{G}\langle\mathcal{C}^*\rangle$ for every isolated isomorphism χ of $\mathcal{G}$ over $\mathcal{F}$.*

Proof. That the condition is necessary is obvious; suppose then that the condition is satisfied. The differential field $\chi\mathcal{G}$, isomorphic with $\mathcal{G}$, also must have this property, so that $\mathcal{G} = \chi^{-1}(\chi\mathcal{G}) \subset (\chi\mathcal{G})\langle\mathcal{C}^*\rangle$, and χ is strong. It follows (Chapter II, Propositions 2 and 5) that every isomorphism of $\mathcal{G}$ over $\mathcal{F}$ is strong, so that $\mathcal{G}$ is strongly normal over $\mathcal{F}$.

PROPOSITION 2. *If $\mathcal{G}$ is strongly normal over $\mathcal{F}$ then $\mathcal{G}$ is normal over $\mathcal{F}$.*

Proof. Let $\mathcal{F}_1$ be a differential field between $\mathcal{F}$ and $\mathcal{G}$, and let α be an element of $\mathcal{G}$ not in $\mathcal{F}_1$. There exists an isomorphism σ of $\mathcal{G}$ over $\mathcal{F}_1$ such that $\sigma\alpha - \alpha \neq 0$, and by hypothesis σ is strong. By Corollary 2 to Proposition 9 of Chapter II there exists an automorphism τ of $\mathcal{G}$ over $\mathcal{F}_1$ which is a specialization of σ such that $\tau\alpha - \alpha \neq 0$. Therefore $\mathcal{G}$ is normal over $\mathcal{F}$.

That $\mathcal{G}$ can be normal over $\mathcal{F}$ without being strongly normal over $\mathcal{F}$ is shown by the example of § 1; using the notation of that example we easily see that for each complex number d there exists a unique isomorphism σ of $\mathcal{G}$ over $\mathcal{F}$ such that $\sigma x = x + d$, $\sigma e^x = e^d e^x$, $\sigma e^{ix} = e^{id} e^{ix}$, $\sigma e^{x^2} = e^{d^2} e^{2dx} e^{x^2}$, and it is obvious that if $2d$ is not a gaussian integer then $e^{2dx} \notin \mathcal{G} = \mathcal{F}\langle x, e^x, e^{ix}, e^{x^2}\rangle$ so that $\mathcal{G}\langle\sigma\mathcal{G}\rangle \subsetneq \mathcal{G}\langle\mathcal{C}_\sigma\rangle = \mathcal{G}$, whence $\mathcal{G}$ is not strongly normal over $\mathcal{F}$.

PROPOSITION 3. *If $\mathcal{G}$ is a strongly normal extension of $\mathcal{F}$ then the group of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ is algebraic.*

Proof. By Proposition 1 of Chapter II there exists a finite number of isomorphisms $\chi_1, \cdots, \chi_h$ of $\mathcal{G}$ over $\mathcal{F}$ such that every isomorphism of $\mathcal{G}$ over $\mathcal{F}$ is a specialization of one of these; by hypothesis χ_i is strong and is

therefore a generic element of an irreducible set $\mathfrak{M}_i$ of automorphisms of $\mathfrak{H}$ over $\mathfrak{F}$. The group of all automorphisms of $\mathfrak{H}$ over $\mathfrak{F}$ is $\bigcup \mathfrak{M}_i$ and therefore is algebraic.

3. The fundamental theorems. Whenever $\mathfrak{H}$ is strongly normal over $\mathfrak{F}$ we shall denote the group of all automorphisms of $\mathfrak{H}$ over $\mathfrak{F}$ by $\mathfrak{G}$, and for each intermediate differential field $\mathfrak{F}_1$ we shall denote the group of all automorphisms of $\mathfrak{H}$ over $\mathfrak{F}_1$ by $\mathfrak{G}(\mathfrak{F}_1)$.

THEOREM 2. *If $\mathfrak{H}$ is strongly normal over $\mathfrak{F}$, then the mapping $\mathfrak{F}_1 \to \mathfrak{G}(\mathfrak{F}_1)$ is one-to-one from the set of all differential fields between $\mathfrak{F}$ and $\mathfrak{H}$ onto the set of all algebraic groups in $\mathfrak{G}$, and has the property that* $\dim \mathfrak{G}(\mathfrak{F}_1) = \partial^0 \mathfrak{H}/\mathfrak{F}_1$.

Proof. By Proposition 2 and Theorem 1 the mapping is one-to-one, and by Proposition 3 $\mathfrak{G}(\mathfrak{F}_1)$ is algebraic; a generic element of a component of $\mathfrak{G}(\mathfrak{F}_1)$ is an isolated isomorphism of $\mathfrak{H}$ over $\mathfrak{F}_1$ whence (Chapter II, Proposition 2) $\dim \mathfrak{G}(\mathfrak{F}_1) = \partial^0 \mathfrak{H}/\mathfrak{F}_1$. It remains to prove the mapping is onto. To this end let $\mathfrak{G}_1$ be any algebraic group in $\mathfrak{G}$, and let $\mathfrak{F}_1$ be the differential field of invariants of $\mathfrak{G}_1$; we shall show that $\mathfrak{G}_1 = \mathfrak{G}(\mathfrak{F}_1)$, thereby proving that $\mathfrak{G}_1$ is algebraic and completing the proof of the theorem. Now, it is obvious that $\mathfrak{G}_1 \subseteq \mathfrak{G}(\mathfrak{F}_1)$. Suppose that $\mathfrak{G}_1 \neq \mathfrak{G}(\mathfrak{F}_1)$. Then, if $\eta_1, \cdots, \eta_n$ are elements such that $\mathfrak{H} = \mathfrak{F}\langle \eta_1, \cdots, \eta_n \rangle$, there exists a differential polynomial in $\mathfrak{H}\{y_1, \cdots, y_n\}$ which vanishes at

$$(\sigma\eta_1, \cdots, \sigma\eta_n) \tag{4}$$

for every $\sigma \in \mathfrak{G}_1$ but not for every $\sigma \in \mathfrak{G}(\mathfrak{F}_1)$; of all such differential polynomials let F be one with a minimal number of terms, and assume without loss of generality that one of the coefficients in F is 1. Let τ be any element of $\mathfrak{G}_1$ and let F_τ be the differential polynomial obtained when each coefficient ϕ in F is replaced by $\tau\phi$. Then

$$F_\tau(\sigma\eta_1, \cdots, \sigma\eta_n) = \tau F(\tau^{-1}\sigma\eta_1, \cdots, \tau^{-1}\sigma\eta_n) = 0$$

for every $\sigma \in \mathfrak{G}_1$, so that $F - F_\tau$ vanishes at (4) for every $\sigma \in \mathfrak{G}_1$. Now $F - F_\tau$ has fewer terms than F has, so that $F - F_\tau$ must vanish at (4) for every $\sigma \in \mathfrak{G}(\mathfrak{F}_1)$. If $F - F_\tau$ were not 0 there would exist an element $\gamma \in \mathfrak{H}$ such that $F - \gamma(F - F_\tau)$ has fewer terms than F has; but $F - \gamma(F - F_\tau)$ obviously vanishes at (4) for every $\sigma \in \mathfrak{G}_1$ but not for every $\sigma \in \mathfrak{G}(\mathfrak{F}_1)$, and therefore can not have fewer terms that F has. It follows that $F - F_\tau = 0$. Since this is true for every $\tau \in \mathfrak{G}_1$, each coefficient in F belongs to $\mathfrak{F}_1$. From

this it follows that F vanishes at (4) for every $\sigma \in \mathfrak{G}(\mathfrak{F}_1)$. This contradiction proves that $\mathfrak{G}_1 = \mathfrak{G}(\mathfrak{F}_1)$, and completes the proof of the theorem.

THEOREM 3. *If $\mathfrak{H}$ is strongly normal over $\mathfrak{F}$, $\mathfrak{F}_1$ is a differential field between $\mathfrak{F}$ and $\mathfrak{H}$, and σ_1 is an isomorphism of $\mathfrak{F}_1$ over $\mathfrak{F}$ into $\mathfrak{H}$, then there exists an automorphism $\sigma \in \mathfrak{G}$ which is an extension of σ_1.*

Proof. σ_1 can be extended to an isomorphism σ' of $\mathfrak{H}$ over $\mathfrak{F}$. By Corollary 2 to Proposition 9 of Chapter II there exist an automorphism σ of $\mathfrak{H}$ which is a specialization of σ' and which therefore coincides with σ_1 on $\mathfrak{F}_1$.

THEOREM 4. *Let $\mathfrak{H}$ be a strongly normal extension of $\mathfrak{F}$ and let $\mathfrak{F}_1$ be a differential field between $\mathfrak{F}$ and $\mathfrak{H}$. Then the following five conditions are equivalent.* 1) *$\mathfrak{F}_1$ is strongly normal over $\mathfrak{F}$*; 2) *$\mathfrak{F}_1$ is normal over $\mathfrak{F}$*; 3) *$\mathfrak{F}_1$ is weakly normal over $\mathfrak{F}$*; 4) *$\sigma\mathfrak{F}_1 \subseteq \mathfrak{F}_1$ for every $\sigma \in \mathfrak{G}$*; 5) *$\mathfrak{G}(\mathfrak{F}_1)$ is a normal subgroup of $\mathfrak{G}$. When these conditions are satisfied then the mapping, which to each $\sigma \in \mathfrak{G}$ assigns the restriction of σ to $\mathfrak{F}_1$, is a homomorphism with kernel $\mathfrak{G}(\mathfrak{F}_1)$ of $\mathfrak{G}$ onto the group of all automorphisms of $\mathfrak{F}_1$ over $\mathfrak{F}$.*

Proof. We already know that 1) implies 2) and that 2) implies 3); also, by Theorem 1, 4) is equivalent to 5). To prove the equivalence of the five conditions it suffices to show that 3) implies 5) and that 4) implies 1). To settle the first point suppose that $\mathfrak{G}(\mathfrak{F}_1)$ is not a normal subgroup of $\mathfrak{G}$, so that the normalisor $N(\mathfrak{G}(\mathfrak{F}_1)) \neq \mathfrak{G}$; by Proposition 17 of Chapter II, $N(\mathfrak{G}(\mathfrak{F}_1))$ is an algebraic group in $\mathfrak{G}$, so that by Theorem 2 there exists a differential field $\mathfrak{F}_2$ between $\mathfrak{F}$ and $\mathfrak{F}_1$ with $\mathfrak{F}_2 \neq \mathfrak{F}$, such that $N(\mathfrak{G}(\mathfrak{F}_1)) = \mathfrak{G}(\mathfrak{F}_2)$. Let α be any element of $\mathfrak{F}_2$ not in $\mathfrak{F}$. If σ_1 is any automorphism of $\mathfrak{F}_1$ over $\mathfrak{F}$ then (Theorem 3) σ_1 can be extended to an automorphism $\sigma \in \mathfrak{G}$; since $\sigma\mathfrak{F}_1 = \sigma_1\mathfrak{F}_1 = \mathfrak{F}_1$, we see that $\tau\sigma\beta = \sigma\beta$ for every $\beta \in \mathfrak{F}_1$ and every $\tau \in \mathfrak{G}(\mathfrak{F}_1)$, that is $\sigma^{-1}\tau\sigma\beta = \beta$, so that $\sigma^{-1}\tau\sigma \in \mathfrak{G}(\mathfrak{F}_1)$, whence $\sigma \in N(\mathfrak{G}(\mathfrak{F}_1)) = \mathfrak{G}(\mathfrak{F}_2)$. Therefore $\sigma_1\alpha = \sigma\alpha = \alpha$, that is, every automorphism of $\mathfrak{F}_1$ over $\mathfrak{F}$ leaves α invariant, so that $\mathfrak{F}_1$ is not weakly normal over $\mathfrak{F}$. Thus, if $\mathfrak{F}_1$ is weakly normal over $\mathfrak{F}$ then $\mathfrak{G}(\mathfrak{F}_1)$ is a normal subgroup of $\mathfrak{G}$.

To settle the second point, suppose that $\mathfrak{F}_1$ is not strongly normal over $\mathfrak{F}$. Then (Proposition 1) there exists an isomorphism σ_1 of $\mathfrak{F}_1$ over $\mathfrak{F}$ such that $\sigma_1\mathfrak{F}_1 \not\subseteq \mathfrak{F}_1\langle\mathfrak{H}^*\rangle$, and σ_1 can be extended to an isomorphism σ of $\mathfrak{H}$ over $\mathfrak{F}$. Let θ be an element of $\sigma_1\mathfrak{F}_1 = \sigma\mathfrak{F}_1$ which does not belong to

$\mathfrak{F}_1\langle\mathfrak{C}^*\rangle$. We claim there exists a $\tau \in \mathfrak{G}(\mathfrak{F}_1)$ such that $\tau\theta \neq \theta$. Indeed, since $\theta \in \sigma\mathfrak{F}_1 \subseteq \mathfrak{H}\langle\sigma\mathfrak{H}\rangle = \mathfrak{H}\langle\mathfrak{C}_\sigma\rangle$, we may write

$$\theta B(c_1, \cdots, c_r) = \sum_{i=0}^{s-1} A_i(c_1, \cdots, c_r)d^i,$$

where $B, A_0, \cdots, A_{s-1}$ are polynomials in $\mathfrak{H}[u_1, \cdots, u_r]$ without common divisor, one of the coefficients in B is 1, $c_1, \cdots, c_r$ are elements of $\mathfrak{C}_\sigma$ algebraically independent over $\mathfrak{C}$ (and hence over $\mathfrak{H}$), and d is an element of $\mathfrak{C}_\sigma$ which is algebraic of some degree s over $\mathfrak{H}\langle c_1, \cdots, c_r\rangle$. If $\tau \in \mathfrak{G}(\mathfrak{F}_1)$ has the property that $\tau\theta = \theta$ then,

$$\sum_{i=0}^{s-1} (B_\tau(c_1, \cdots, c_r)A_i(c_1, \cdots, c_r) - B(c_1, \cdots, c_r)A_{i\tau}(c_1, \cdots, c_r))\, d^i = 0$$

(where in general for any polynomial C we denote by C_τ the polynomial obtained upon replacing each coefficient in C by its image under τ), whence $B_\tau A_i = BA_{i\tau}$ $(0 \leq i \leq s-1)$. Because $B, A_0, \cdots, A_{s-1}$ have no common divisor and one of the coefficients in B is 1, it follows that $B_\tau = B$, $A_{i\tau} = A_i$ $(0 \leq i \leq s-1)$. Therefore if θ were invariant under every $\tau \in \mathfrak{G}(\mathfrak{F}_1)$ then so would every coefficient in B and each A_i, and these coefficients would all belong to $\mathfrak{F}_1$, contradicting the fact that $\theta \notin \mathfrak{F}_1\langle\mathfrak{C}^*\rangle$. This establishes our claim that for some $\tau \in \mathfrak{G}(\mathfrak{F}_1)$ we have $\tau\theta \neq \theta$.

Now $\theta \in \sigma\mathfrak{F}_1$, so that there exists an element $\zeta \in \mathfrak{F}_1$ such that $\theta = \sigma\zeta$. For this ζ and the above τ we have $\tau\sigma\zeta \neq \sigma\zeta$. By Chapter II, Proposition 9, there exists an automorphism σ_0 of $\mathfrak{H}$ which is a specialization of σ for which $\tau\sigma_0\zeta \neq \sigma_0\zeta$. Since $\tau \in \mathfrak{G}(\mathfrak{F}_1)$ this means that $\sigma_0\zeta \notin \mathfrak{F}_1$ and since $\zeta \in \mathfrak{F}_1$ this shows that $\sigma_0\mathfrak{F}_1 \not\subseteq \mathfrak{F}_1$. We have thus shown that if $\mathfrak{F}_1$ is not strongly normal over $\mathfrak{F}$ then $\sigma_0\mathfrak{F}_1 \not\subseteq \mathfrak{F}_1$ for some $\sigma_0 \in \mathfrak{G}$, so that 4) implies 1). This completes the proof that the five conditions are equivalent.

When these conditions are satisfied then (Theorem 1) the mapping which to each automorphism in $\mathfrak{G}$ assigns its restriction to $\mathfrak{F}_1$ is a homomorphism with kernel $\mathfrak{G}(\mathfrak{F}_1)$ of $\mathfrak{G}$ *into* the group of all automorphisms of $\mathfrak{F}_1$ over $\mathfrak{F}$. That this homomorphism is *onto* follows from Theorem 3.

COROLLARY. *If $\mathfrak{H}$ and $\mathfrak{K}$ are strongly normal extension of $\mathfrak{F}$ such that the field of constants of $\mathfrak{F}\langle\mathfrak{H}, \mathfrak{K}\rangle$ is $\mathfrak{C}$, then $\mathfrak{F}\langle\mathfrak{H}, \mathfrak{K}\rangle$ and $\mathfrak{H} \cap \mathfrak{K}$ are strongly normal over $\mathfrak{F}$.*

Proof. Let σ be any isomorphism of $\mathfrak{F}\langle\mathfrak{H}, \mathfrak{K}\rangle$ over $\mathfrak{F}$; the restrictions of σ to $\mathfrak{H}$ and to $\mathfrak{K}$ are isomorphisms of $\mathfrak{H}$ and of $\mathfrak{K}$ over $\mathfrak{F}$ and, since $\mathfrak{H}$ and $\mathfrak{K}$ are strongly normal over $\mathfrak{F}$, we have

$$\sigma(\mathfrak{F}\langle\mathfrak{H}, \mathfrak{K}\rangle) = \mathfrak{F}\langle\sigma\mathfrak{H}, \sigma\mathfrak{K}\rangle$$
$$\subseteq \mathfrak{F}\langle\mathfrak{H}\langle\mathfrak{C}^*\rangle, \mathfrak{K}\langle\mathfrak{C}^*\rangle\rangle = \mathfrak{F}\langle\mathfrak{H}, \mathfrak{K}\rangle\langle\mathfrak{C}^*\rangle.$$

Therefore (Proposition 1) $\mathfrak{F}\langle\mathfrak{G},\mathfrak{H}\rangle$ is strongly normal over $\mathfrak{F}$. Now let τ be any automorphism of $\mathfrak{F}\langle\mathfrak{G},\mathfrak{H}\rangle$ over $\mathfrak{F}$. Since $\mathfrak{G}$ and $\mathfrak{H}$ are both strongly normal over $\mathfrak{F}$ we see by Theorem 4 that $\tau\mathfrak{G}\subseteq\mathfrak{G}$ and $\tau\mathfrak{H}\subseteq\mathfrak{H}$, whence $\tau(\mathfrak{G}\cap\mathfrak{H})\subseteq\mathfrak{G}\cap\mathfrak{H}$, so that (Theorem 4) $\mathfrak{G}\cap\mathfrak{H}$ is strongly normal over $\mathfrak{F}$.

THEOREM 5. *Let* $\mathfrak{m}$ *be a set of elements such that the field of constants of* $\mathfrak{G}\langle\mathfrak{m}\rangle$ *is* $\mathfrak{k}$, *let* $\mathfrak{F}^\dagger=\mathfrak{F}\langle\mathfrak{m}\rangle$ *and* $\mathfrak{G}^\dagger=\mathfrak{G}\langle\mathfrak{m}\rangle$, *and denote the group of all automorphisms of* $\mathfrak{G}^\dagger$ *over* $\mathfrak{F}^\dagger$ *by* $\mathfrak{G}^\dagger$. *If* $\mathfrak{G}$ *is strongly normal over* $\mathfrak{F}$ *then* $\mathfrak{G}^\dagger$ *is strongly normal over* $\mathfrak{F}^\dagger$, *and the mapping which to each* $\sigma^\dagger\in\mathfrak{G}^\dagger$ *assigns the restriction* σ *of* $\sigma^\dagger$ *to* $\mathfrak{G}$ *is an isomorphism of* $\mathfrak{G}^\dagger$ *onto* $\mathfrak{G}(\mathfrak{F}^\dagger\cap\mathfrak{G})$ *which maps every algebraic group* $\mathfrak{M}^\dagger$ *in* $\mathfrak{G}^\dagger$ *onto an algebraic group* $\mathfrak{M}$ *in* $\mathfrak{G}$ *of the same dimension*; *if* $\mathfrak{M}^\dagger$ *is irreducible then so is* $\mathfrak{M}$.

Proof. Let $\rho^\dagger$ be any isomorphism of $\mathfrak{G}^\dagger$ over $\mathfrak{F}^\dagger$ and let ρ be the restriction of $\rho^\dagger$ to $\mathfrak{G}$. Then ρ is an isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$, and

$$\mathfrak{G}^\dagger\langle\rho^\dagger\mathfrak{G}^\dagger\rangle=\mathfrak{G}\langle\mathfrak{m}\rangle\langle(\rho\mathfrak{G})\langle\mathfrak{m}\rangle\rangle=\mathfrak{G}\langle\mathfrak{k}_\rho\rangle\langle\mathfrak{m}\rangle=\mathfrak{G}^\dagger\langle\mathfrak{k}_\rho\rangle;$$

it follows (Proposition 1) that $\mathfrak{G}^\dagger$ is strongly normal over $\mathfrak{F}^\dagger$, and also (Corollary 5 to Proposition 3 of Chapter 1) that the field of constants of $\mathfrak{G}^\dagger\langle\rho^\dagger\mathfrak{G}^\dagger\rangle$ is $\mathfrak{k}_\rho$.

The mapping $\sigma^\dagger\to\sigma$ is obviously a homomorphism of $\mathfrak{G}^\dagger$ into $\mathfrak{G}$. If σ is the identity automorphism of $\mathfrak{G}$ then $\sigma^\dagger$ is obviously the identity automorphism of $\mathfrak{G}^\dagger$; therefore this homomorphism is an isomorphism of $\mathfrak{G}^\dagger$ onto some subgroup $\mathfrak{G}_1$ of $\mathfrak{G}$. An element $\alpha\in\mathfrak{G}$ is invariant under every $\sigma\in\mathfrak{G}_1$, that is under every $\sigma^\dagger\in\mathfrak{G}^\dagger$, if and only if $\alpha\in\mathfrak{F}^\dagger$; it follows that $\mathfrak{G}(\mathfrak{F}^\dagger\cap\mathfrak{G})$ is the smallest algebraic group in $\mathfrak{G}$ containing $\mathfrak{G}_1$, and therefore (Chapter II, Proposition 14) also the smallest algebraic set containg $\mathfrak{G}_1$.

Now let $\mathfrak{M}^\dagger$ be any irreducible set in $\mathfrak{G}^\dagger$, let $\rho^\dagger$ be a generic element of $\mathfrak{M}^\dagger$, and let ρ be the restriction of $\rho^\dagger$ to $\mathfrak{G}$. Let $\mathfrak{M}$ be the set of all restrictions to $\mathfrak{G}$ of elements of $\mathfrak{M}^\dagger$, and let $\mathfrak{M}_0$ be the irreducible set in $\mathfrak{G}$ with generic element ρ. Because every element of $\mathfrak{M}^\dagger$ is a specialization of $\rho^\dagger$, every element of $\mathfrak{M}$ is a specialization of ρ, so that $\mathfrak{M}\subseteq\mathfrak{M}_0$. If $\gamma_1,\cdots,\gamma_q$ are constants such that $\mathfrak{k}_\rho=\mathfrak{k}\langle\gamma_1,\cdots,\gamma_q\rangle$ then, by Proposition 9 of Chapter II and the fact that the field of constants of $\mathfrak{G}^\dagger\langle\rho^\dagger\mathfrak{G}^\dagger\rangle$ is $\mathfrak{k}_\rho$, we know that there exists a polynomial $M\in\mathfrak{k}[u_1,\cdots,u_q]$ with $M(\gamma_1,\cdots,\gamma_q)\neq0$ such that whenever $c_1,\cdots,c_q$ are constants with $M(c_1,\cdots,c_q)\neq0$ then there exists a unique isomorphism $\tau^\dagger$ of $\mathfrak{G}^\dagger$ over $\mathfrak{F}^\dagger$ for which $((\tau^\dagger\alpha)_{\alpha\in\mathfrak{G}^\dagger},c_1,\cdots,c_q)$ is a specialization of $((\rho^\dagger\alpha)_{\alpha\in\mathfrak{G}^\dagger},\gamma_1,\cdots,\gamma_q)$ over $\mathfrak{G}^\dagger$. But it is easy to see that if $\eta_1,\cdots,\eta_n$ are elements of $\mathfrak{G}$ such that $\mathfrak{G}=\mathfrak{F}\langle\eta_1,\cdots,\eta_n\rangle$, then there exists a differential polynomial

$K \in \mathfrak{F}\{y_1, \cdots, y_n\}$ with $K(\rho\eta_1, \cdots, \rho\eta_n) \neq 0$ which has the following property: whenever τ is an isomorphism of $\mathfrak{F}$ over $\mathfrak{F}$ which is a specialization of ρ with $K(\tau\eta_1, \cdots, \tau\eta_n) \neq 0$ then there exist unique constants $c_1, \cdots, c_q$ with $M(c_1, \cdots, c_q) \neq 0$ such that $((\tau\alpha)_{\alpha \in \mathfrak{G}}, c_1, \cdots, c_q)$ is a specialization of $((\rho\alpha)_{\alpha \in \mathfrak{G}}, \gamma_1, \cdots, \gamma_q)$ over $\mathfrak{F}$. It follows that every specialization τ of ρ such that $K(\tau\eta_1, \cdots, \tau\eta_n) \neq 0$ is the restriction to $\mathfrak{F}$ of a specialization $\tau^\dagger$ of $\rho^\dagger$, and that $\tau^\dagger$ is an automorphism of $\mathfrak{F}^\dagger$ if τ is an automorphism of $\mathfrak{F}$. From this it is not difficult to conclude that $\mathfrak{M}_0 - \mathfrak{M}$ is contained in a finite union of irreducible sets in $\mathfrak{G}$ of lower dimension that $\mathfrak{M}_0$.

Suppose now that, in addition to being an irreducible set in $\mathfrak{G}^\dagger$, $\mathfrak{M}^\dagger$ is also a group; then $\mathfrak{M}$ is a group in $\mathfrak{G}$. If we let $\mathfrak{N}$ denote the intersection of all the algebraic sets in $\mathfrak{G}$ which contain $\mathfrak{M}$ then $\mathfrak{M} \subseteq \mathfrak{N} \subseteq \mathfrak{M}_0$. By Chapter I, Proposition 14, $\mathfrak{N}$ is an algebraic group. Also, the irreducible set $\mathfrak{M}_0$ is the union of $\mathfrak{N}$ and a finite union of irreducible sets of lower dimension than $\mathfrak{M}_0$, so that $\mathfrak{M}_0 = \mathfrak{N}$, $\mathfrak{M}_0$ is an algebraic group in $\mathfrak{G}$, and (Chapter II, Proposition 15) $\mathfrak{M}_0 = \mathfrak{M}$. Therefore $\mathfrak{M}$ is an irreducible algebraic group in $\mathfrak{G}$ of dimension

$$= \partial^0 \mathfrak{F}\langle \rho\mathfrak{F} \rangle / \mathfrak{F} = \partial^0 \mathfrak{L}_\rho / \mathfrak{L} = \partial^0 \mathfrak{F}^\dagger \langle \rho^\dagger \mathfrak{F}^\dagger \rangle / \mathfrak{F}^\dagger = \dim \mathfrak{M}^\dagger.$$

Thus every irreducible algebraic group in $\mathfrak{G}^\dagger$ is mapped onto an irreducible algebraic group in $\mathfrak{G}$ of the same dimension. Since every algebraic group is the union of the finite number of cosets of its component of the identity, a similar remark holds for not necessarily irreducible algebraic groups. In particular, the image of $\mathfrak{G}^\dagger$ is $\mathfrak{G}_1 = \mathfrak{G}(\mathfrak{F}^\dagger \cap \mathfrak{F})$.

COROLLARY. *Let $\mathfrak{F}$ be strongly normal over $\mathfrak{F}$, and let $\mathfrak{F}_0, \mathfrak{F}_1, \cdots, \mathfrak{F}_r$ be differential fields such that*

$$\mathfrak{F} = \mathfrak{F}_0 \subseteq \mathfrak{F}_1 \subseteq \cdots \subseteq \mathfrak{F}_r, \quad \mathfrak{F} \subseteq \mathfrak{F}_r,$$

and $\mathfrak{F}_i$ is a strongly normal extension of $\mathfrak{F}_{i-1}$ $(1 \leq i \leq r)$; denote the group of all automorphisms of $\mathfrak{F}_i$ over $\mathfrak{F}_{i-1}$ by $\mathfrak{G}_i$ $(1 \leq i \leq r)$. Then

$$(5) \qquad \mathfrak{G} = \mathfrak{G}(\mathfrak{F}_0 \cap \mathfrak{F}) \supseteq \mathfrak{G}(\mathfrak{F}_1 \cap \mathfrak{F}) \supseteq \cdots \supseteq \mathfrak{G}(\mathfrak{F}_r \cap \mathfrak{F}) = \{\iota\}$$

is a normal chain of subgroups of $\mathfrak{G}$, $\mathfrak{G}_i(\mathfrak{F}_i \cap \mathfrak{F}\langle \mathfrak{F}_{i-1} \rangle)$ is a normal subgroup of $\mathfrak{G}_i$, and for $i = 1, 2, \cdots, r$

$$(6) \qquad \mathfrak{G}(\mathfrak{F}_{i-1} \cap \mathfrak{F}) / \mathfrak{G}(\mathfrak{F}_i \cap \mathfrak{F}) \approx \mathfrak{G}_i / \mathfrak{G}_i(\mathfrak{F}_i \cap \mathfrak{F}\langle \mathfrak{F}_{i-1} \rangle),$$

$$(7) \qquad \dim \mathfrak{G}(\mathfrak{F}_{i-1} \cap \mathfrak{F}) - \dim \mathfrak{G}(\mathfrak{F}_i \cap \mathfrak{F}) = \dim \mathfrak{G}_i - \dim \mathfrak{G}_i(\mathfrak{F}_i \cap \mathfrak{F}\langle \mathfrak{F}_{i-1} \rangle).$$

Proof. If $r = 1$ the assertions follow immediately from Theorem 4. Let $r > 1$ and suppose the corollary proved for lower values of r.

By the corollary to Theorem 4, $\mathfrak{H}_1 \cap \mathfrak{H}$ is strongly normal over $\mathfrak{F}$ so that (Theorem 4) $\mathfrak{G}(\mathfrak{H}_1 \cap \mathfrak{H})$ is a normal subgroup $\mathfrak{G} = \mathfrak{G}(\mathfrak{H}_0 \cap \mathfrak{H})$ and $\mathfrak{G}_1(\mathfrak{H}_1 \cap \mathfrak{H})$ is a normal subgroup of $\mathfrak{G}_1$. The two factor groups $\mathfrak{G}/\mathfrak{G}(\mathfrak{H}_1 \cap \mathfrak{H})$ and $\mathfrak{G}_1/\mathfrak{G}_1(\mathfrak{H}_1 \cap \mathfrak{H})$ are isomorphic to the group of all automorphisms of $\mathfrak{H}_1 \cap \mathfrak{H}$ over $\mathfrak{F}$, and therefore to each other, so that (6) holds for $i = 1$. By Theorem 2, also (7) holds for $i = 1$, as in that case both members equal $\partial^0(\mathfrak{H}_1 \cap \mathfrak{H})/\mathfrak{F}$.

To complete the proof we consider the group $\mathfrak{G}^\dagger$ of all automorphisms of $\mathfrak{H}\langle\mathfrak{H}_1\rangle$ over $\mathfrak{F}\langle\mathfrak{H}_1\rangle = \mathfrak{H}_1$. By the theorem $\mathfrak{H}\langle\mathfrak{H}_1\rangle$ is strongly normal over $\mathfrak{H}_1$. By the corollary (case $r - 1$)

$$(8) \qquad \begin{aligned} \mathfrak{G}^\dagger &= \mathfrak{G}^\dagger(\mathfrak{H}_1 \cap \mathfrak{H}\langle\mathfrak{H}_1\rangle) \\ &\supseteq \mathfrak{G}^\dagger(\mathfrak{H}_2 \cap \mathfrak{H}\langle\mathfrak{H}_1\rangle) \supseteq \cdots \supseteq \mathfrak{G}^\dagger(\mathfrak{H}_r \cap \mathfrak{H}\langle\mathfrak{H}_1\rangle) = \{\iota\} \end{aligned}$$

is a normal chain of subgroups of $\mathfrak{G}^\dagger$, $\mathfrak{G}_i(\mathfrak{H}_i \cap \mathfrak{H}\langle\mathfrak{H}_{i-1}\rangle)$ is a normal subgroup of $\mathfrak{G}_i$ $(2 \leq i \leq r)$, and

$$\mathfrak{G}^\dagger(\mathfrak{H}_{i-1} \cap \mathfrak{H}\langle\mathfrak{H}_1\rangle)/\mathfrak{G}^\dagger(\mathfrak{H}_i \cap \mathfrak{H}\langle\mathfrak{H}_1\rangle) \approx \mathfrak{G}_i/\mathfrak{G}_i(\mathfrak{H}_i \cap \mathfrak{H}\langle\mathfrak{H}_{i-1}\rangle),$$

$$\begin{aligned} \dim \mathfrak{G}^\dagger(\mathfrak{H}_{i-1} \cap \mathfrak{H}\langle\mathfrak{H}_1\rangle) &- \dim \mathfrak{G}^\dagger(\mathfrak{H}_i \cap \mathfrak{H}\langle\mathfrak{H}_1\rangle) \\ &= \dim \mathfrak{G}_i - \dim \mathfrak{G}_i(\mathfrak{H}_i \cap \mathfrak{H}\langle\mathfrak{H}_{i-1}\rangle) \end{aligned}$$

for $i = 2, \cdots, r$. By the theorem, the mapping $\sigma^\dagger \to \sigma$ which assigns to each $\sigma^\dagger \in \mathfrak{G}^\dagger$ the restriction σ of $\sigma^\dagger$ to $\mathfrak{H}$ is a dimension-preserving isomorphism which maps the normal chain (8) onto the normal chain

$$\mathfrak{G}(\mathfrak{H}_1 \cap \mathfrak{H}) \supseteq \mathfrak{G}(\mathfrak{H}_2 \cap \mathfrak{H}) \supseteq \cdots \supseteq \mathfrak{G}(\mathfrak{H}_r \cap \mathfrak{H}) = \{\iota\}.$$

It now quickly follows that (6) and (7) hold for $2 \leq i \leq r$, and the proof of the corollary is complete.

4. Primitive elements. An element α will be called *primitive* over $\mathfrak{F}$ if $\delta_i\alpha \in \mathfrak{F}$ $(1 \leq i \leq m)$. It is obvious that if α is primitive over $\mathfrak{F}$ with $\delta_i\alpha = a_i$ $(1 \leq i \leq m)$, and if β is primitive over $\mathfrak{F}$ with $\delta_i\beta = b_i$ $(1 \leq i \leq m)$, and if we set $\eta = \alpha + \beta$, then η is primitive over $\mathfrak{F}$ with $\delta_i\eta = a_i + b_i$ $(1 \leq i \leq m)$.

Let α be primitive over $\mathfrak{F}$ and suppose that the field of constants of $\mathfrak{F}\langle\alpha\rangle$ is $\mathfrak{k}$. For every isomorphism σ of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$

$$\delta_i(\sigma\alpha - \alpha) = \sigma\delta_i\alpha - \delta_i\alpha = \delta_i\alpha - \delta_i\alpha = 0 \qquad (1 \leq i \leq m),$$

so that $c(\sigma) = \sigma\alpha - \alpha$ is a constant. Because $\sigma\alpha = \alpha + c(\sigma)$ $\mathfrak{F}\langle\alpha\rangle$ is strongly normal over $\mathfrak{F}$ and every isomorphism of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ is strong. If σ_1, σ_2 are two isomorphisms of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ then

$$\sigma_1\sigma_2\alpha = \sigma_1(\alpha + c(\sigma_2)) = \alpha + c(\sigma_1) + c(\sigma_2),$$

so that $c(\sigma_1\sigma_2) = c(\sigma_1) + c(\sigma_2)$. Since $c(\sigma) = 0$ only when σ is the identity automorphism ι of $\mathfrak{F}\langle\alpha\rangle$, it follows that $\sigma \to c(\sigma)$ is an isomorphism of the group of all automorphisms of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ into $\mathfrak{C}^+$ (the additive group of $\mathfrak{C}$). If $\alpha \in \mathfrak{F}$ the automorphism group consists solely of ι, and the corresponding group in $\mathfrak{C}^+$ is the zero group. Suppose $\alpha \notin \mathfrak{F}$. Then there exists an automorphism σ of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ different from ι, and therefore with $c(\sigma) \neq 0$. Now if c is a constant then $\alpha + c$ is a generic specialization of α over $\mathfrak{F}$ if and only if $\alpha + c$ is a specialization of α over $\mathfrak{F}$, that is if and only if $\alpha + c$ is a zero of every differential polynomial in $\mathfrak{F}\{y\}$ which vanishes at α, and this takes place if and only if c is a zero of a certain ideal of polynomials in $\mathfrak{C}[u]$. Since this ideal of polynomials has infinitely many zeros, namely $c(\sigma^n) = nc(\sigma)$ for every integer n, it is the zero ideal, so that $\alpha + c$ is a generic specialization of α over $\mathfrak{F}$ and $c = c(\sigma)$ for a suitable isomorphism σ of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ for every constant c. Summarizing: *Let α be primitive over $\mathfrak{F}$ and the field of constants of $\mathfrak{F}\langle\alpha\rangle$ be $\mathfrak{C}$. Then $\mathfrak{F}\langle\alpha\rangle$ is strongly normal over $\mathfrak{F}$; either $\alpha \in \mathfrak{F}$, or else α is transcendental over $\mathfrak{F}$ and the mapping $\sigma \to \sigma\alpha - \alpha$ is an isomorphism of the group of all automorphisms of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ onto $\mathfrak{C}^+$ and there exists no differential field between $\mathfrak{F}$ and $\mathfrak{F}\langle\alpha\rangle$ other than $\mathfrak{F}$ and $\mathfrak{F}\langle\alpha\rangle$.*

We note the trivial fact that, with the law of composition $(c_1, c_2) \to c_1 + c_2$, one-dimensional affine space becomes a group variety $\boldsymbol{D}$ in the sense of Weil [10], and that $\mathfrak{C}^+$ is the subgroup of $\boldsymbol{D}$ consisting of all points of $\boldsymbol{D}$ which are rational over $\mathfrak{C}$.

5. Exponential elements. An element α will be called *exponential* over $\mathfrak{F}$ if $\alpha \neq 0$ and $\alpha^{-1}\delta_i\alpha \in \mathfrak{F}$ $(1 \leq i \leq m)$. It is obvious that if α is exponential over $\mathfrak{F}$ with $\alpha^{-1}\delta_i\alpha = a_i$ $(1 \leq i \leq m)$, and if β is exponential over $\mathfrak{F}$ with $\beta^{-1}\delta_i\beta = b_i$ $(1 \leq i \leq m)$, and if we set $\eta = \alpha\beta$, the η is exponential over $\mathfrak{F}$ with $\eta^{-1}\delta_i\eta = a_i + b_i$ $(1 \leq i \leq m)$.

Let α be exponential over $\mathfrak{F}$ and suppose that the field of constants of $\mathfrak{F}\langle\alpha\rangle$ is $\mathfrak{C}$. For every isomorphism σ of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$

$$\begin{aligned}\delta_i(\alpha^{-1}\sigma\alpha) &= \alpha^{-2}(\alpha \cdot \delta_i\sigma\alpha - \sigma\alpha \cdot \delta_i\alpha) \\ &= \alpha^{-2}(\alpha \cdot \sigma\alpha \cdot \sigma(\alpha^{-1}\delta_i\alpha) - \sigma\alpha \cdot \alpha \cdot \alpha^{-1}\delta_i\alpha) = 0 \qquad (1 \leq i \leq m),\end{aligned}$$

so that $c(\sigma) = \alpha^{-1}\sigma\alpha$ is a constant. Because $\sigma\alpha = c(\sigma)\alpha$, $\mathfrak{F}\langle\alpha\rangle$ is strongly normal over $\mathfrak{F}$ and every isomorphism of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ is strong. If σ_1, σ_2 are two isomorphism of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ then

$$\sigma_1\sigma_2\alpha = \sigma_1(c(\sigma_2)\alpha) = c(\sigma_1)c(\sigma_2)\alpha,$$

so that $c(\sigma_1\sigma_2) = c(\sigma_1)c(\sigma_2)$. Since $c(\sigma) = 1$ only when $\sigma = \iota$, it follows that $\sigma \to c(\sigma)$ is an isomorphism of the group of all automorphisms of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ into $\mathfrak{C}^\times$ (the multiplicative group of nonzero elements of $\mathfrak{C}$). If α is algebraic over $\mathfrak{F}$ then the automorphism group is finite, say of order d; since every finite subgroup of $\mathfrak{C}^\times$ is cyclic the automorphism group is cyclic, being generated by a single automorphism, say σ. It follows that $c(\sigma)$ is a primitive d-th root of unity, so that $\sigma(\alpha^d) = (c(\sigma)\alpha)^d = \alpha^d$, whence $\alpha^d \in \mathfrak{F}$. If α is transcendental over $\mathfrak{F}$ then the automorphism group is infinite, and it follows, much as in the case of primitive elements, that $c\alpha$ is a generic specialization of α over $\mathfrak{F}$ for every nonzero constant c, so that $\sigma \to c(\sigma)$ is an isomorphism of the automorphism group of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ onto $\mathfrak{C}^\times$; furthermore, if $\mathfrak{F}_1$ is a differential field between $\mathfrak{F}$ and $\mathfrak{F}\langle\alpha\rangle$ then, since α is exponential over $\mathfrak{F}_1$, the mapping $\sigma \to c(\sigma)$ maps the group of all automorphisms of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}_1$ either onto a cyclic group of some finite order d (in which case $\alpha^d \in \mathfrak{F}_1$, whence $\mathfrak{F}_1 = \mathfrak{F}\langle\alpha^d\rangle$) or else onto the whole group $\mathfrak{C}^\times$ (in which case $\mathfrak{F}_1 = \mathfrak{F}$). Summarizing: *Let α be exponential over $\mathfrak{F}$ and the field of constants of $\mathfrak{F}\langle\alpha\rangle$ be $\mathfrak{C}$. Then $\mathfrak{F}\langle\alpha\rangle$ is strongly normal over $\mathfrak{F}$; either there exists an integer $d > 0$ such that $\alpha^d \in \mathfrak{F}$, or else α is transcendental over $\mathfrak{F}$ and the mapping $\sigma \to \alpha^{-1}\sigma\alpha$ is an isomorphism of the group of all automorphisms of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ onto $\mathfrak{C}^\times$ and the only differential fields between $\mathfrak{F}$ and $\mathfrak{F}\langle\alpha\rangle$ are those of the form $\mathfrak{F}\langle\alpha^d\rangle$, where d is an integer ≥ 0.*

We note the trivial fact that, with the law of composition $(c_1, c_2) \to c_1c_2$, one-dimensional affine space with the origin deleted becomes a group variety $\boldsymbol{E}$, and that $\mathfrak{C}^\times$ is the subgroup of $\boldsymbol{E}$ consisting of all points of $\boldsymbol{E}$ which are rational over $\mathfrak{C}$.

6. Weierstrassian elements. An element α will be called *weierstrassian* over $\mathfrak{F}$ if α is not a constant and there exist two elements $g_2, g_3 \in \mathfrak{C}$ with $27g_3^2 - g_2^3 \neq 0$ and m elements $a_1, \cdots, a_m \in \mathfrak{F}$ such that

$$(\delta_i\alpha)^2 = a_i^2(4\alpha^3 - g_2\alpha - g_3) \qquad (1 \leq i \leq m).$$

The condition $27g_3^2 - g_2^3 \neq 0$ is equivalent to the condition that the polynomial $4y^3 - g_2y - g_3$ have no multiple root; the constants g_2, g_3, which are

uniquely determined if α is transcendental over $\mathcal{F}$, will be called the *invariants* of α. Since any a_i may obviously be replaced by $-a_i$ *we may suppose (and in the future we always shall suppose, without expressly mentioning the fact) that* $(\delta_1\alpha : \cdots : \delta_m\alpha) = (a_1 : \cdots : a_m)$.

In order to study weierstrassian elements with invariants g_2, g_3 we consider the irreducible algebraic curve in the projective plane defined by the equation

$$X_0X_2^2 - 4X_1^3 + g_2X_0^2X_1 + g_3X_0^3 = 0.$$

This curve, which is of genus 1, has precisely one point on the line at infinity $X_0 = 0$, namely the point $(0:0:1)$. On this curve there exists a law of composition, which we shall write multiplicatively, with respect to which the points of the curve form a group, the unity element being the point of infinity; the curve is, in the language of Weil [10], a group variety (actually an abelian variety) of dimension 1. If $(1:\xi_1:\xi_2)$ and $(1:\eta_1:\eta_2)$ are two points of this curve and if $\eta_1 \neq \xi_1$, then their product is given by the formulae

$$(9)\quad \begin{cases} (1:\xi_1:\xi_2)(1:\eta_1:\eta_2) = (1:\zeta_1:\zeta_2), \\ \zeta_1 = -\xi_1 - \eta_1 + \frac{1}{4}\left(\dfrac{\xi_2 - \eta_2}{\xi_1 - \eta_1}\right)^2, \\ \zeta_2 = -\frac{1}{2}(\xi_2 + \eta_2) + \frac{1}{2}(\xi_1 + \eta_1)\dfrac{\xi_2 - \eta_2}{\xi_1 - \eta_1} - \frac{1}{4}\left(\dfrac{\xi_2 - \eta_2}{\xi_1 - \eta_1}\right)^3; \end{cases}$$

the inverse of any point $(1:\xi_1:\xi_2)$ of the curve is the point $(1:\xi:-\xi_2)$, so that there are precisely three points of order 2, namely the points $(1:e_1:0)$, $(1:e_2:0)$, $(1:e_3:0)$, where e_1, e_2, e_3 are the roots of $4y^3 - g_2y - g_3$; the square of any point $(1:\xi_1:\xi_2)$ with $\xi_2 \neq 0$ is given by the formulae

$$(10)\quad \begin{cases} (1:\xi_1:\xi_2)^2 = (1:\zeta_1:\zeta_2), \\ \zeta_1 = -2\xi_1 + 4^{-1}(6\xi_1^2 - \frac{1}{2}g_2)^2\xi_2^{-2}, \\ \zeta_2 = -\xi_2 + 3\xi_1(6\xi_1^2 - \frac{1}{2}g_2)\xi_2^{-1} - 4^{-1}(6\xi_1^2 - \frac{1}{2}g_2)^3\xi_2^{-3}. \end{cases}$$

These facts are well-known and are not difficult to verify directly. We shall denote this group variety by $\boldsymbol{W}(g_2, g_3)$.

If α is weierstrassian over $\mathcal{F}$ with $(\delta_i\alpha)^2 = a_i^2(4\alpha^3 - g_2\alpha - g_3)$ $(1 \leq i \leq m)$ then, since α is not a constant, $\delta_i\alpha \neq 0$ for at least one value of i and $a_i \neq 0$ whenever $\delta_i\alpha \neq 0$. By the convention made above, the point $(1:\alpha:a_i^{-1}\delta_i\alpha)$ does not depend on the choice of i from among those values for which $a_i \neq 0$; this point obviously belongs to $\boldsymbol{W}(g_2, g_3)$.

LEMMA 2. *Let α be weierstrassian over $\mathfrak{F}$ with*

$$(\delta_i\alpha)^2 = a_i^2(4\alpha^3 - g_2\alpha - g_3) \qquad (1 \leq i \leq m)$$

and $a_{i_0} \neq 0$, let β be weierstrassian over $\mathfrak{F}$ with

$$(\delta_i\beta)^2 = b_i^2(4\beta^3 - g_2\beta - g_3) \qquad (1 \leq i \leq m)$$

and $b_{j_0} \neq 0$, suppose that

$$(1 : \alpha : a_{i_0}^{-1}\delta_{i_0}\alpha)\,(1 : \beta : b_{j_0}^{-1}\delta_{j_0}\beta) \neq (0 : 0 : 1),$$

and let

$$(1 : \alpha : a_{i_0}^{-1}\delta_{i_0}\alpha)\,(1 : \beta : b_{j_0}^{-1}\delta_{j_0}\beta) = (1 : \eta : \zeta).$$

Then $\delta_i\eta = (a_i + b_i)\zeta$ $(1 \leq i \leq m)$, so that either η and ζ are both constants or else η is weierstrassian over $\mathfrak{F}$ with

$$(\delta_i\eta)^2 = (a_i + b_i)^2(4\eta^3 - g_2\eta - g_3) \qquad (1 \leq i \leq m).$$

Proof. Suppose first that $\alpha \neq \beta$. Then by (9) and a simple computation we find that

$$\begin{aligned}(11)\quad \zeta = {} & (\alpha - \beta)^{-3}(-\beta^2(3\alpha + \beta) + \tfrac{1}{4}g_2(\alpha + 3\beta) + g_3)\,a_{i_0}^{-1}\delta_{i_0}\alpha \\ & + (\alpha - \beta)^{-3}(\alpha^2(\alpha + 3\beta) - \tfrac{1}{4}g_2(3\alpha + \beta) - g_3)\,b_{j_0}^{-1}\delta_{j_0}\beta.\end{aligned}$$

On the other hand

$$\delta_i\alpha = a_i a_{i_0}^{-1}\delta_{i_0}\alpha, \qquad \delta_i\beta = b_i b_{j_0}^{-1}\delta_{j_0}\beta,$$

$$\delta_i(a_{i_0}^{-1}\delta_{i_0}\alpha) = (6\alpha^2 - \tfrac{1}{2}g_2)a_i, \qquad \delta_i(b_{j_0}^{-1}\delta_{j_0}\beta) = (6\beta^2 - \tfrac{1}{2}g_2)b_i,$$

so that from (9) we find that

$$\begin{aligned}\delta_i\eta = {} & -\delta_i\alpha - \delta_i\beta + \tfrac{1}{2}\,\frac{a_{i_0}^{-1}\delta_{i_0}\alpha - b_{j_0}^{-1}\delta_{j_0}\beta}{\alpha - \beta}\left(\frac{\delta_i(a_{i_0}^{-1}\delta_{i_0}\alpha - b_{j_0}^{-1}\delta_{j_0}\beta)}{\alpha - \beta}\right. \\ & \left. - \frac{(a_{i_0}^{-1}\delta_{i_0}\alpha - b_{j_0}\delta_{j_0}\beta)\,(\delta_i\alpha - \delta_i\beta)}{(\alpha - \beta)^2}\right) \\ = {} & -a_i a_{i_0}^{-1}\delta_{i_0}\alpha - b_i b_{j_0}^{-1}\delta_{j_0}\beta \\ & + \tfrac{1}{2}\,\frac{a_{i_0}^{-1}\delta_{i_0}\alpha - b_{j_0}^{-1}\delta_{j_0}\beta}{\alpha - \beta}\;\frac{(6\alpha^2 - \tfrac{1}{2}g_2)a_i - (6\beta^2 - \tfrac{1}{2}g_2)b_i}{\alpha - \beta} \\ & - \tfrac{1}{2}\,\frac{(a_{i_0}^{-1}\delta_{i_0}\alpha - b_{j_0}^{-1}\delta_{j_0}\beta)^2\,(a_i a_{i_0}^{-1}\delta_{i_0}\alpha - b_i b_{j_0}^{-1}\delta_{j_0}\beta)}{(\alpha - \beta)^3}.\end{aligned}$$

The coefficient of a_i here is easily seen to be the second member of (11), and likewise for the coefficient of b_i here. It follows that $\delta_i\eta = (a_i + b_i)\zeta$.

Now suppose that $\alpha = \beta$. Because by hypothesis

$$(1 : \alpha : a_{i_0}^{-1}\delta_{i_0}\alpha)\,(1 : \beta : b_{j_0}^{-1}\delta_{j_0}\beta) \neq (0 : 0 : 1),$$

we have $a_i = b_i$ $(1 \leq i \leq m)$. Therefore (10) is applicable in computing $(1 : \eta : \zeta)$, and we find on the one hand

$$\begin{aligned}\zeta &= -a_{i_0}^{-1}\delta_{i_0}\alpha + 3\alpha(6\alpha^2 - \tfrac{1}{2}g_2)(a_{i_0}^{-1}\delta_{i_0}\alpha)^{-1} \\ &\quad - \tfrac{1}{4}(6\alpha^2 - \tfrac{1}{2}g_2)^3(a_{i_0}^{-1}\delta_{i_0}\alpha)^{-3} \\ &= (-1 + 3\alpha(6\alpha^2 - \tfrac{1}{2}g_2)(4\alpha^3 - g_2\alpha - g_3)^{-1} \\ &\quad - \tfrac{1}{4}(6\alpha^2 - \tfrac{1}{2}g_2)^3(4\alpha^3 - g_2\alpha - g_3)^{-2})a_{i_0}^{-1}\delta_{i_0}\alpha,\end{aligned}$$

and on the other hand

$$\eta = -2\alpha + \tfrac{1}{4}(6\alpha^2 - \tfrac{1}{2}g_2)^2(4\alpha^3 - g_2\alpha - g_3)^{-1},$$

so that

$$\begin{aligned}\delta_i\eta &= (-2 + 6\alpha(6\alpha^2 - \tfrac{1}{2}g_2)(4\alpha^3 - g_2\alpha - g_3)^{-1} \\ &\quad - \tfrac{1}{2}(6\alpha^2 - \tfrac{1}{2}g_2)^3(4\alpha^3 - g_2\alpha - g_3)^{-2})\delta_i\alpha,\end{aligned}$$

whence $\delta_i\eta = 2a_i\zeta = (a_i + b_i)\zeta$, q. e. d.

Now let α be an element which is weierstrassian over $\mathfrak{F}$ with

$$(\delta_i\alpha)^2 = a_i^2(4\alpha^3 - g_2\alpha - g_3) \qquad (1 \leq i \leq m),$$

and suppose that the field of constants of $\mathfrak{F}\langle\alpha\rangle$ is $\mathfrak{k}$. We let i_0 denote any subscript such that $a_{i_0} \neq 0$. For any isomorphism σ of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$,

$$\begin{aligned}&(1 : \sigma\alpha : a_{i_0}^{-1}\delta_{i_0}\sigma\alpha)(1 : \alpha : -a_{i_0}^{-1}\delta_{i_0}\alpha) \\ &\qquad = (1 : \sigma\alpha : a_{i_0}^{-1}\delta_{i_0}\sigma\alpha)(1 : \alpha : a_{i_0}^{-1}\delta_{i_0}\alpha)^{-1}\end{aligned}$$

is a point of the group variety $\mathfrak{W}(g_2, g_3)$; we denote this point by $P(\sigma)$. If $\sigma = \iota$ then obviously $P(\sigma) = (0 : 0 : 1)$. Suppose $\sigma \neq \iota$; then $P(\sigma) \neq (0 : 0 : 1)$, and we may write $P(\sigma) = (1 : c_1(\sigma) : c_2(\sigma))$. It follows from Lemma 2 that $\delta_i c_1(\sigma) = (a_i - a_i)c_2(\sigma) = 0$ $(1 \leq i \leq m)$, so that $c_1(\sigma)$ and, therefore, $c_2(\sigma)$ are constants. Therefore for every isomorphism σ of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ we have

$$(1 : \sigma\alpha : a_{i_0}^{-1}\delta_{i_0}\sigma\alpha) = P(\sigma)(1 : \alpha : a_{i_0}^{-1}\delta_{i_0}\alpha)$$

and $\sigma\alpha \in \mathfrak{F}\langle\alpha\rangle\langle\mathfrak{k}^*\rangle$, so that $\mathfrak{F}\langle\alpha\rangle$ is strongly normal over $\mathfrak{F}$ and σ is strong. If σ_1, σ_2 are two isomorphisms of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ then

$$\begin{aligned}&(1 : \sigma_1\sigma_2\alpha : a_{i_0}^{-1}\delta_{i_0}\sigma_1\sigma_2\alpha) \\ &\qquad = P(\sigma_2)(1 : \sigma_1\alpha : c_{i_0}^{-1}\delta_{i_0}\sigma_1\alpha) = P(\sigma_2)P(\sigma_1)(1 : \alpha : a_{i_0}^{-1}\delta_{i_0}\alpha),\end{aligned}$$

so that $P(\sigma_1\sigma_2) = P(\sigma_1)P(\sigma_2)$. Since $P(\sigma) = (0 : 0 : 1)$ only when $\sigma = \iota$, it follows that $\sigma \to P(\sigma)$ is an isomorphism of the group of all automorphisms

of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ into the group $\mathfrak{W}(g_2, g_3; \mathfrak{C})$ consisting of all points of $\mathfrak{W}(g_2, g_3)$ which are rational over $\mathfrak{C}$.

Let c_1, c_2 be any two constants such that $(1 : c_1 : c_2) \in \mathfrak{W}(g_2, g_3)$ and set $(1 : \beta : \beta') = (1 : c_1 : c_2)(1 : \alpha : a_{i_0}^{-1}\delta_{i_0}\alpha)$. Now if (β, β') is a specialization of $(\alpha, a_{i_0}^{-1}\delta_{i_0}\alpha)$ over $\mathfrak{F}$ then the specialization is generic, and this will be the case if and only if $\beta = \sigma\alpha$, $\beta' = a_{i_0}^{-1}\delta_{i_0}\sigma\alpha$ for some isomorphism σ of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$, that is, if and only if $(1 : c_1 : c_2) = P(\sigma)$ for some such σ. On the other hand (β, β') is a specialization of $(\alpha, a_{i_0}^{-1}\delta_{i_0}\alpha)$ over $\mathfrak{F}$ if and only if (c_1, c_2) is a zero of a certain set of polynomials with coefficients in $\mathfrak{F}\langle\alpha\rangle$, and therefore if and only if (c_1, c_2) is a zero of a certain set of polynomials with coefficients in $\mathfrak{C}$. It follows that $\sigma \to P(\sigma)$ maps the group of all automorphisms of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ onto the intersection with $\mathfrak{W}(g_2, g_3; \mathfrak{C})$ of a subgroup of $\mathfrak{W}(g_2, g_3)$ which is a subvariety (not necessarily irreducible) of $\mathfrak{W}(g_2, g_3)$. Of course the subvarieties of an irreducible curve other than the curve itself are finite.

We may summarize these facts as follows: *Let α be weierstrassian over $\mathfrak{F}$ and the field of constants of $\mathfrak{F}\langle\alpha\rangle$ be $\mathfrak{C}$. Then $\mathfrak{F}\langle\alpha\rangle$ is strongly normal over $\mathfrak{F}$; either α is algebraic over $\mathfrak{F}$ and the mapping*

$$\sigma \to (1 : \alpha : a_{i_0}^{-1}\delta_{i_0}\alpha)^{-1}(1 : \sigma\alpha : a_{i_0}^{-1}\delta_{i_0}\sigma\alpha)$$

is an isomorphism of the group of all automorphisms of $\mathfrak{F}\langle\alpha\rangle$ over $\mathfrak{F}$ onto a finite subgroup of $\mathfrak{W}(g_2, g_3; \mathfrak{C})$, or else α is transcendental over $\mathfrak{F}$ and this mapping is an isomorphism onto $\mathfrak{W}(g_2, g_3; \mathfrak{C})$. It can be shown, although we do not do so here, that if $\mathfrak{F}_1$ is a differential field between $\mathfrak{F}$ and $\mathfrak{F}\langle\alpha\rangle$ other than $\mathfrak{F}$ then $\mathfrak{F}_1$ contains an element α_1 weierstrassian over $\mathfrak{F}$ such that $\mathfrak{F}_1 = \mathfrak{F}\langle\alpha_1\rangle$.

7. Picard-Vessiot extensions. Let $\mathfrak{G}$ be a Picard-Vessiot extension of $\mathfrak{F}$. Then (Kolchin [3] and [6]), for suitable generators $\eta_1, \cdots, \eta_n$ of $\mathfrak{G}$ over $\mathfrak{F}$, every isomorphism σ of $\mathfrak{G}$ over $\mathfrak{F}$ satisfies equations

$$\sigma\eta_j = \sum_{i=1}^{n} c_{ij}(\sigma)\eta_i \qquad (1 \leq j \leq n), \tag{12}$$

where each $c_{ij}(\sigma)$ is a constant, and these equations establish a one-to-one correspondence between the set of all isomorphisms σ of $\mathfrak{G}$ over $\mathfrak{F}$ and a certain set of invertible matrices $(c_{ij}(\sigma))$ of degree n with constant coordinates. It follows that $\mathfrak{G}$ is strongly normal over $\mathfrak{F}$, and each isomorphism of $\mathfrak{G}$ over $\mathfrak{F}$ is strong. The mapping $\sigma \to (c_{ij}(\sigma))$ is an isomorphism of the algebraic group $\mathfrak{G}$ of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ onto a certain algebraic

matric group $\mathfrak{G}_M$ over $\mathscr{C}$; algebraic subgroups of $\mathfrak{G}$ are mapped thereby onto algebraic subgroups of $\mathfrak{G}_M$ of the same dimension. If σ, τ are isomorphisms of $\mathscr{G}$ over $\mathscr{F}$ then τ is a specialization of σ if and only if $(c_{ij}(\tau))$ is a specialization of $(c_{ij}(\sigma))$ over $\mathscr{C}$.

We shall say that a differential field $\mathscr{H}$ is an extension of $\mathscr{F}$ by algebraic, primitive, exponential and weierstrassian elements if $\mathscr{H}$ contains a finite family of elements $\alpha_1, \cdots, \alpha_r$ such that $\mathscr{H} = \mathscr{F}\langle\alpha_1, \cdots, \alpha_r\rangle$ and for each i $(1 \leq i \leq r)$ α_i is either algebraic, or primitive, or exponential, or weierstrassian over $\mathscr{F}\langle\alpha_1, \cdots, \alpha_{i-1}\rangle$. If $\mathscr{H} = \mathscr{F}\langle\alpha_1, \cdots, \alpha_r\rangle$ and for each i $(1 \leq i \leq r)$ α_i is either algebraic, or primitive, or exponential over $\mathscr{F}\langle\alpha_1, \cdots, \alpha_{i-1}\rangle$, and if the field of constants of $\mathscr{H}$ is $\mathscr{C}$, then $\mathscr{H}$ is called a *liouvillian* extension of $\mathscr{F}$.

THEOREM 6. *If a Picard-Vessiot extension of $\mathscr{F}$ is contained in an extension of $\mathscr{F}$ by algebraic, primitive, exponential, and weierstrassian elements with field of constants $\mathscr{C}$,*[8] *then the Picard-Vessiot extension is a liouvillian extension of $\mathscr{F}$.*

Proof. By the hypothesis, the results of the preceding three sections, and the corollary to Theorem 5, the group of all automorphisms of the Picard-Vessiot extension of $\mathscr{F}$ has a normal chain of algebraic subgroups in which each factor group is either finite or abelian. Therefore (Kolchin [3], § 8, Theorem 1) the component of the identity of this group of automorphisms is solvable. It follows that the Picard-Vessiot extension is a liouvillian extension.

8. Extensions of transcendence degree 1; formulation of the theorem. In §§ 4-6 we saw that if α is transcendental and either primitive or exponential or weierstrassian over $\mathscr{F}$, and if the field of constants of $\mathscr{F}\langle\alpha\rangle$ is $\mathscr{C}$, then $\mathscr{F}\langle\alpha\rangle$ is a strongly normal extension of $\mathscr{F}$ of transcendence degree 1. We shall now state a theorem which implies that every strongly normal (indeed, every weakly normal) extension of $\mathscr{F}$ of transcendence degree 1 can be obtained by combining the adjunction of an element of one of these three types with algebraic adjunctions.

Let $\mathscr{G}$ be a weakly normal extension of $\mathscr{F}$. It is a simple matter to see that the relative algebraic closure $\mathscr{F}^0$ of $\mathscr{F}$ in $\mathscr{G}$ is a normal algebraic extension of $\mathscr{F}$ (in the classical sense) of finite degree. If the group $\mathfrak{G}$ of all automorphisms of $\mathscr{G}$ over $\mathscr{F}$ is finite then $\mathscr{G} = \mathscr{F}^0$; therefore if $\mathscr{G}$ is trans-

[8] By considerations similar to those of Kolchin [4] it can be shown that this restriction on the field of constants of the extension may be omitted.

cendental over $\mathfrak{F}$ then $\mathfrak{G}$ is infinite, and it easily follows that the group $\mathfrak{G}^0$ of all automorphisms of $\mathfrak{H}$ over $\mathfrak{F}^0$ is infinite. We therefore may state our theorem in the following form.

THEOREM 7. *Let $\mathfrak{H}$ be of transcendence degree 1 over $\mathfrak{F}$, let $\mathfrak{F}$ be relatively algebraically closed in $\mathfrak{H}$, and let the group $\mathfrak{G}$ of all automorphisms of $\mathfrak{H}$ over $\mathfrak{F}$ be infinite. Then there exists an element $\alpha \in \mathfrak{H}$ such that either α is primitive over $\mathfrak{F}$ and $\mathfrak{H} = \mathfrak{F}\langle\alpha\rangle$, or α is exponential over $\mathfrak{F}$ and $\mathfrak{H} = \mathfrak{F}\langle\alpha\rangle$, or α is weierstrassian over $\mathfrak{F}$ and $\mathfrak{H}$ is an algebraic*[9] *extension of $\mathfrak{F}\langle\alpha\rangle$. In the last case, if $\mathfrak{F}$ is algebraically closed then the weierstrassian element α may be chosen so that $\mathfrak{H} = \mathfrak{F}\langle\alpha\rangle$.*

This theorem will be proved in §§ 9-11.

An immediate consequence of Theorem 7, the results of §§ 4-6, and the Corollary of Theorem 5, is the following.

COROLLARY. *Let $\mathfrak{H}$ be any strongly normal extension of $\mathfrak{F}$. If $\mathfrak{H}$ is contained in an extension of $\mathfrak{F}$ by algebraic, primitive, exponential, and weierstrassian elements then the group $\mathfrak{G}$ of all automorphisms of $\mathfrak{H}$ over $\mathfrak{F}$ has a normal chain $\mathfrak{G} = \mathfrak{G}_0 \supseteq \cdots \supseteq \mathfrak{G}_s = \{\iota\}$ of algebraic groups such that $\dim \mathfrak{G}_{i-1} - \dim \mathfrak{G}_i \leq 1$ $(1 \leq i \leq s)$. Conversely, if $\mathfrak{G}$ has such a normal chain then $\mathfrak{H}$ is itself an extension of $\mathfrak{F}$ by algebraic, primitive, exponential, and weierstrassian elements.*

9. The proof begun: reduction to the case of algebraically closed ground field. We shall show in this section that if Theorem 7 holds when $\mathfrak{F}$ is algebraically closed then it holds in general.

Let $\mathfrak{F}^\dagger$ be the algebraic closure of $\mathfrak{F}$ and let $\mathfrak{H}^\dagger = \mathfrak{H}\langle\mathfrak{F}^\dagger\rangle$. Since $\mathfrak{F}$ is relatively algebraically closed in $\mathfrak{H}$, the degree of each element of $\mathfrak{F}^\dagger$ over $\mathfrak{F}$ equals its degree over $\mathfrak{H}$. Now every $P \in \mathfrak{F}^\dagger\{y_1, \cdots, y_n\}$ can be written in the form $P = \sum_{i=0}^{d-1} P_i \lambda^i$, where each $P_i \in \mathfrak{F}\{y_1, \cdots, y_n\}$ and λ is an element of $\mathfrak{F}^\dagger$ of some degree d over $\mathfrak{F}$, and therefore over $\mathfrak{H}$. Consequently a family $(\eta_1, \cdots, \eta_n)$ of elements of $\mathfrak{H}$ is a zero of P if and only if it is a zero of $P_0, P_1, \cdots, P_{d-1}$. From this it follows that every automorphism of $\mathfrak{H}$ over $\mathfrak{F}$ can be extended to an automorphism of $\mathfrak{H}^\dagger$ over $\mathfrak{F}^\dagger$. From the hypothesis of Theorem 7 it therefore follows that there are infinitely many automorphisms of $\mathfrak{H}^\dagger$ over $\mathfrak{F}^\dagger$. By the assumption that the theorem holds when the ground

[9] Abelian.

field is algebraically closed we conclude that there exists an element η such that $\mathfrak{G}^\dagger = \mathfrak{F}^\dagger\langle\eta\rangle$ and η is either primitive or exponential or weierstrassian over $\mathfrak{F}^\dagger$. Thus there exist elements $a_1, \cdots, a_m \in \mathfrak{F}^\dagger$ such that either $\delta_i\eta = a_i$ $(1 \leq i \leq m)$ or $\eta^{-1}\delta_i\eta = a_i$ $(1 \leq i \leq m)$ or $(4\eta^3 - g_2\eta - g_3)^{-1}(\delta\eta)^2 = a_i^2$ $(1 \leq i \leq m)$, where in the last case $g_2, g_3 \in \mathfrak{C}$, $27{g_3}^2 - {g_2}^3 \neq 0$, and $(\delta_1\eta : \cdots : \delta_m\eta) = (a_1 : \cdots : a_m)$.

Let σ be any automorphism of $\mathfrak{G}^\dagger$ over $\mathfrak{F}^\dagger$ other than the identity. We may write, in the respective cases, $\sigma\eta = \eta + c$ or $\sigma\eta = c\eta$ or $(1 : \sigma\eta : a_i^{-1}\delta_i\sigma\eta) = (1 : c_1 : c_2)(1 : \eta : a_i^{-1}\delta_i\eta)$, where in the first case $c \in \mathfrak{C}^+$, in the second case $c \in \mathfrak{C}^\times$, and in the third case $(1 : c_1 : c_2) \in \mathfrak{W}(g_2, g_3; \mathfrak{C})$ and $a_i \neq 0$. Now let τ be any automorphism of $\mathfrak{G}^\dagger$ over $\mathfrak{G}$. It is clear that $\tau\mathfrak{F}^\dagger = \mathfrak{F}^\dagger$; therefore $\sigma\tau\theta = \tau\theta = \tau\sigma\theta$ for all $\theta \in \mathfrak{F}^\dagger$. If σ happens to be one of the infinitely many automorphisms of $\mathfrak{G}^\dagger$ over $\mathfrak{F}^\dagger$ which are extensions of automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ then $\sigma\tau\theta = \sigma\theta = \tau\sigma\theta$ for every $\theta \in \mathfrak{G}$ whence, since $\mathfrak{G}^\dagger = \mathfrak{G}\langle\mathfrak{F}^\dagger\rangle$, $\sigma\tau = \tau\sigma$. Because $\mathfrak{G}^\dagger = \mathfrak{F}^\dagger\langle\eta\rangle$ we may write $\tau\eta = f(\eta)$, where $f \in \mathfrak{F}^\dagger\langle y\rangle$, and for arbitrary σ we shall have $\sigma\tau = \tau\sigma$ if and only if $\sigma\tau\eta = \tau\sigma\eta$, that is $f(\eta + c) = f(\eta) + c$ in the first case, $f(c\eta) = cf(\eta)$ in the second case, and

$$f\left(-\eta - c_1 + \tfrac{1}{4}\left(\frac{a_i^{-1}\delta_i\eta - c_2}{\eta - c_1}\right)^2\right) = -f(\eta) - c_1 + \tfrac{1}{4}\left(\frac{a_i^{-1}\delta_i f(\eta) - c_2}{f(\eta) - c_1}\right)^2$$

in the third case. Since this condition is satisfied for infinitely many choices of $c \in \mathfrak{C}^+$ in the first case, of $c \in \mathfrak{C}^\times$ in the second case, and of $(1 : c_1 : c_2) \in \mathfrak{W}(g_2, g_3; \mathfrak{C})$ in the third case, it must be satisfied identically. Thus σ commutes with every automorphism of $\mathfrak{G}^\dagger$ over $\mathfrak{G}$.

Let $\tau_1, \cdots, \tau_n$ be automorphisms of $\mathfrak{G}^\dagger$ over $\mathfrak{G}$ such that the restrictions of $\tau_1, \cdots, \tau_n$ to $\mathfrak{G}\langle\eta\rangle$ are distinct and constitute the set of all isomorphisms of $\mathfrak{G}\langle\eta\rangle$ over $\mathfrak{G}$ (so that n equals the degree of $\mathfrak{G}\langle\eta\rangle$ over $\mathfrak{G}$). Since $\sigma\tau_j = \tau_j\sigma$, we have, in the respective cases, $\sigma\tau_j\eta = \tau_j\eta + c$, or $\sigma\tau_j\eta = c\tau_j\eta$, or

$$(1 : \sigma\tau_j\eta : \sigma\tau_j(a_i^{-1}\delta_i\eta)) = (1 : c_1 : c_2)(1 : \tau_j\eta : \tau_j(a_i^{-1}\delta_i\eta)).$$

We now consider the first case. Letting $\alpha = \sum_j \tau_j\eta$, we see that $\alpha \in \mathfrak{G}$; also, $\delta_i\alpha = \sum_j \tau_j a_i \in \mathfrak{F}$, so that α is primitive over $\mathfrak{F}$. Furthermore, $\sigma\alpha = \alpha + nc$, so that $\alpha \notin \mathfrak{F}^\dagger$; it follows (§ 4) that $\mathfrak{G}^\dagger = \mathfrak{F}^\dagger\langle\eta\rangle = \mathfrak{F}^\dagger\langle\alpha\rangle$. If θ is any element of $\mathfrak{G}$ then we may write $\theta = \sum \phi_i\alpha^i / \sum \psi_i\alpha^i$, where $\sum \phi_i y^i$ and $\sum \psi_i y^i$ are relatively prime polynomials in $\mathfrak{F}^\dagger[y]$ such that the leading coefficient in $\sum \psi_i y^i$ is 1. For any automorphism τ of $\mathfrak{G}^\dagger$ over $\mathfrak{G}$ we have $\tau\alpha = \alpha$ and $\tau\theta = \theta$, so that $\sum \tau\phi_i\alpha^i / \sum \tau\psi_i\alpha^i = \sum \phi_i\alpha^i / \sum \psi_i\alpha^i$, whence $(\sum \tau\phi_i\alpha^i)(\sum \psi_i\alpha^i) = (\sum \phi_i\alpha^i)(\sum \tau\psi_i\alpha^i)$. Because of the relative primeness

mentioned above and the fact that $\sum \psi_i y^i$ has leading coefficient 1 we conclude that $\sum \tau\phi_i \alpha^i = \sum \phi_i \alpha^i$ and $\sum \tau\psi_i \alpha^i = \sum \psi_i \alpha^i$, so that $\tau\phi_i = \phi_i$ and $\tau\psi_i = \psi_i$ for all i. Since τ is any automorphism of $\mathcal{G}^\dagger$ over $\mathcal{G}$ it follows that ϕ_i and ψ_i belong to $\mathcal{G} \cap \mathcal{F}^\dagger = \mathcal{F}$, so that $\theta \in \mathcal{F}\langle\alpha\rangle$. Thus $\mathcal{G} = \mathcal{F}\langle\alpha\rangle$, and the reduction is complete in the first case.

We turn to the second case. We assert that there exist an integer $r > 0$ and a nonzero element $\phi \in \mathcal{F}^\dagger$ such that $(\phi\eta)^r \in \mathcal{G}$. Indeed, $\sigma \prod \tau_j\eta = c^n \prod \tau_j\eta$, so that if we set $\chi = \eta^{-n} \prod \tau_j\eta$ then $\sigma\chi = \chi$, whence $\chi \in \mathcal{F}^\dagger$; letting ψ be an element of $\mathcal{F}^\dagger$ such that $\psi^n = \chi$, we find that $(\psi\eta)^n = \prod \tau_j\eta \in \mathcal{G}$, which proves our assertion. Of all pairs r, ϕ as above let us suppose we have chosen one for which r is as small as possible. Let $\alpha = (\phi\eta)^r$, so that $\alpha \in \mathcal{G}$; α is exponential over $\mathcal{F}^\dagger$, so that $\alpha^{-1}\delta_i\alpha \in \mathcal{F}^\dagger \cap \mathcal{G} = \mathcal{F}$ $(1 \leq i \leq m)$, whence α is exponential over $\mathcal{F}$. We shall show that $\mathcal{G} = \mathcal{F}\langle\alpha\rangle$, thereby completing the reduction in the second case. Indeed, if θ is any element of $\mathcal{G}$ then $\theta \in \mathcal{G}^\dagger = \mathcal{F}^\dagger\langle\eta\rangle = \mathcal{F}^\dagger\langle\phi\eta\rangle$, so that we may write $\theta = \sum \phi_i(\phi\eta)^i / \sum \psi_i(\phi\eta)^i$, where $\sum \phi_i y^i$ and $\sum \psi_i y^i$ are relatively prime polynomials in $\mathcal{F}^\dagger[y]$ and one of the coefficients ϕ_0, ψ_0 is 1. Since $(\phi\eta)^r \in \mathcal{G}$, for any automorphism τ of $\mathcal{G}^\dagger$ over $\mathcal{G}$ we may write $\tau(\phi\eta) = e\phi\eta$ where e is some r-th root of 1. As $\tau\theta = \theta$, we have $\sum \tau\phi_i \cdot e^i(\phi\eta)^i / \sum \tau\psi_i \cdot e^i(\phi\eta)^i = \sum \phi_i(\phi\eta)^i / \sum \psi_i(\phi\eta)^i$, so that $(\sum \tau\phi_i \cdot e^i(\phi\eta)^i)(\sum \psi_i(\phi\eta)^i) = (\sum \phi_i(\phi\eta)^i)(\sum \tau\psi_i \cdot e^i(\phi\eta)^i)$. Because of the relative primeness mentioned above and the fact that ϕ_0 or ψ_0 is 1, we conclude that $\sum \tau\phi_i \cdot e^i(\phi\eta)^i = \sum \phi_i(\phi\eta)^i$, $\sum \tau\psi_i \cdot e^i(\phi\eta)^i = \sum \psi_i(\phi\eta)^i$, so that $\tau\phi_i = e^{-i}\phi_i$, $\tau\psi_i = e^{-i}\psi_i$. Consider any value of i which is not divisible by r; writing $i = qr + r'$, where $0 < r' < r$, we find that $\tau(\phi_i(\phi\eta)^{r'}) = e^{-i}\phi_i e^{r'}(\phi\eta)^{r'} = \phi_i(\phi\eta)^{r'}$. Since τ is any automorphism of $\mathcal{G}^\dagger$ over $\mathcal{G}$ this implies that $\phi_i(\phi\eta)^{r'} \in \mathcal{G}$. Letting ϕ' be an element of $\mathcal{F}^\dagger$ such that $\phi'^{r'} = \phi_i\phi^{r'}$, we see that $(\phi'\eta)^{r'} \in \mathcal{G}$; because of the minimal nature of r and the relation $0 < r' < r$, we conclude that $\phi' = 0$, whence $\phi_i = 0$. Similarly $\psi_i = 0$ for all i not divisible by r. On the other hand, if i is divisible by r then $\tau\phi_i = e^{-i}\phi_i = \phi_i$, so that $\phi_i \in \mathcal{G} \cap \mathcal{F}^\dagger = \mathcal{F}$, and similarly, $\psi_i \in \mathcal{F}$. It follows that $\theta \in \mathcal{F}\langle(\phi\eta)^r\rangle = \mathcal{F}\langle\alpha\rangle$, so that $\mathcal{G} = \mathcal{F}\langle\alpha\rangle$.

Finally we consider the third case. It is apparent that the point $(1 : \eta : a_i^{-1}\delta_i\eta)^{-n} \prod (1 : \tau_j\eta : \tau_j(a_i^{-1}\delta_i\eta))$ of $W(g_2, g_3)$ is invariant under σ. Since σ is any automorhpism of $\mathcal{G}^\dagger$ over $\mathcal{F}^\dagger$ other than the identity, this point is rational over $\mathcal{F}^\dagger$. Because $\mathcal{F}^\dagger$ is algebraically closed and $W(g_2, g_3)$ is a complete curve in the projective plane, this point can be written in the form P^n, where P is a point of $W(g_2, g_3)$ which is rational over $\mathcal{F}^\dagger$. For this P we have $(1 : \eta : a_i^{-1}\delta_i\eta)^n P^n = \prod (1 : \tau_j\eta : \tau_j(a_i^{-1}\delta_i\eta))$, which is clearly rational over $\mathcal{G}$. Now η is transcendental over $\mathcal{F}^\dagger$, so that $(1 : \eta : a_i^{-1}\delta_i\eta)$ is

a generic point of the curve $\boldsymbol{W}(g_2, g_3)$ over $\mathfrak{F}^\dagger$; since P is rational over $\mathfrak{F}^\dagger$ $(1:\eta:a_i^{-1}\delta_i\eta)P$ is also a generic point of $\boldsymbol{W}(g_2, g_3)$ over $\mathfrak{F}^\dagger$, whence $(1:\eta:a_i^{-1}\delta_i\eta)^n P^n$ is, too. Since the field of constants of $\mathfrak{G}^\dagger$ is clearly $\mathfrak{C}$, which is contained in $\mathfrak{F}^\dagger$, it follows from Lemma 2 (§ 6) that $(1:\eta:a_i^{-1}\delta_i\eta)^n P^n = (1:\alpha:\beta)$, where α is weierstrassian over $\mathfrak{F}^\dagger$ with invariants g_2, g_3. But by the above, $\alpha, \beta \in \mathfrak{G}$, so that α is weierstrassian over $\mathfrak{G} \cap \mathfrak{F}^\dagger = \mathfrak{F}$. Also, α is transcendental over $\mathfrak{F}$, so that $\mathfrak{G}$ is algebraic over $\mathfrak{F}\langle\alpha\rangle$. This completes the reduction in the third and final case.

10. The proof continued: case of genus 0. We assume now, in addition to the hypothesis of Theorem 7, that $\mathfrak{F}$ is algebraically closed. Since $\mathfrak{G}$ is of transcendence degree 1 over $\mathfrak{F}$, we may regard $\mathfrak{G}$ as a field of algebraic functions of one variable over $\mathfrak{F}$; furthermore, every automorphism of the differential field $\mathfrak{G}$ over $\mathfrak{F}$ is obviously an automorphism of the algebraic function field $\mathfrak{G}$. It is a well-known theorem that if the genus of such an algebraic function field is greater than 1 then the group of its automorphisms is finite (for a proof in the general (abstract) case see Iwasawa and Tamagawa [7]). It follows that the algebraic function field $\mathfrak{G}$ has genus 0 to 1. In the present section we treat the case of genus 0 and show that in this case there exists an element α, which is either primitive or exponential over $\mathfrak{F}$, such that $\mathfrak{G} = \mathfrak{F}\langle\alpha\rangle$. In the next section we shall treat the case of genus 1.

Assuming then that $\mathfrak{G}$ has genus 0, we see that $\mathfrak{G}$ is a purely transcendental extension of $\mathfrak{F}$ (see for example Chevalley [2], Chapter II, § 2), that is, that there exists a single element θ transcendental over $\mathfrak{F}$ such that $\mathfrak{G} = \mathfrak{F}(\theta)$. Since $\delta_i\theta \in \mathfrak{G}$ for each i, there exist polynomials $P_1, \cdots, P_m$, $Q \in \mathfrak{F}[y]$ with $Q \neq 0$ such that

$$\delta_i\theta = Q(\theta)^{-1}P_i(\theta), \qquad 1 \leq i \leq m. \tag{13}$$

It is obvious that θ is not a constant, so that $P_i \neq 0$ for at least one value of i. If $c \in \mathfrak{C}$ is not a zero of Q nor of any nonzero P_i, and if $k = \max(\deg P_1, \cdots, \deg P_m, \deg Q)$, then $\bar{Q}(y) = Q(y^{-1} + c)y^k$ and the nonzero expression $\bar{P}_i(y) = -P_i(y^{-1} + c)y^k$ are polynomials of degree k; but if we let $\bar{\theta} = (\theta - c)^{-1}$, so that $\mathfrak{G} = \mathfrak{F}(\theta)$, then $\delta_i\bar{\theta} = \bar{Q}(\bar{\theta})^{-1}\bar{P}_i(\bar{\theta})\bar{\theta}^2$ $(1 \leq i \leq m)$. Therefore we lose no generality in assuming that $\deg P_i = 2 + \deg Q$ for all values of i such that $P_i \neq 0$; we assume too, as we obviously may, that $P_1, \cdots, P_m, Q$ have no common divisor and that the leading coefficient in Q is 1. We denote the degree of Q by d.

Every automorphism of a simple transcendental extension is given by a fractional linear substitution. Therefore if σ is any automorphism of the

differential field $\mathcal{G}$ over $\mathcal{F}$ then there exist elements $a_{11}, a_{12}, a_{21}, a_{22}$ in $\mathcal{F}$ such that

$$\sigma\theta = (a_{21}\theta + a_{22})^{-1}(a_{11}\theta + a_{12}), \qquad |\sigma| \neq 0,$$

where $|\sigma| = a_{11}a_{22} - a_{12}a_{21}$. Applying σ to each side of (13) we find

$$\begin{aligned}(a_{11}\delta_i\theta + \delta_i a_{11}\cdot\theta + \delta_i a_{12})(a_{21}\theta + a_{22})^{-1}&\\ - (a_{11}\theta + a_{12})(a_{21}\theta + a_{22})^{-2}(a_{21}\delta_i\theta + \delta_i a_{21}\cdot\theta + \delta_i a_{22})&\\ = Q((a_{21}\theta + a_{22})^{-1}(a_{11}\theta + a_{12}))^{-1}P_i((a_{21}\theta + a_{22})^{-1}(a_{11}\theta + a_{12})),&\end{aligned}$$

which we rewrite (using (13)) in the form

$$\begin{aligned}(14)\quad & Q((a_{21}\theta + a_{22})^{-1}(a_{11}\theta + a_{12}))(a_{21}\theta + a_{22})^d\cdot(|\sigma| P_i(\theta) + A_i(\theta)Q(\theta))\\ & = Q(\theta)P_i((a_{21}\theta + a_{22})^{-1}(a_{11}\theta + a_{12}))(a_{21}\theta + a_{22})^{d+2},\end{aligned}$$

where

$$\begin{aligned}(15)\quad A_i(y) = {} & (a_{21}\delta_i a_{11} - a_{11}\delta_i a_{21})y^2 + (a_{22}\delta_i a_{11} - a_{11}\delta_i a_{22}\\ & + a_{21}\delta_i a_{12} - a_{12}\delta_i a_{21})y + a_{22}\delta_i a_{12} - a_{12}\delta_i a_{22}.\end{aligned}$$

It follows from (14) that for each i

$$Q((a_{21}y + a_{22})^{-1}(a_{11}y + a_{12}))(a_{21}y + a_{22})^d\cdot|\sigma| P_i(y)$$

is divisible by $Q(y)$, whence the polynomial

$$Q((a_{21}y + a_{22})^{-1}(a_{11}y + a_{12}))(a_{21}y + a_{22})^d$$

is so divisible. It follows that the fractional linear transformation

$$(16)\qquad x \to (a_{21}x + a_{22})^{-1}(a_{11}x + a_{12})$$

permutes the zeros of Q. Since this must happen for each of the infinitely many automorphisms σ of $\mathcal{G}$ over $\mathcal{F}$, and since a fractional linear transformation is uniquely determined by its values at three points, Q can have no more than two distinct zeros.

Supose Q has two distinct zeros ξ_1, ξ_2, so that $Q = (y - \xi_1)^{h_1}(y - \xi_2)^{h_2}$, where $h_1 + h_2 = d$; we suppose for the sake of definiteness that $h_1 \leq h_2$. For each σ in the group of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ the transformation (16) permutes ξ_1, ξ_2. The subgroup of all automorphisms σ for which (16) leaves ξ_1 and ξ_2 invariant is obviously of index 2, and is therefore infinite. For every σ of this infinite subgroup we have $(a_{21}\xi_j + a_{22})^{-1}(a_{11}\xi_j + a_{12}) = \xi_j$, that is $a_{21}\xi_j^2 + (a_{22} - a_{11})\xi_j - a_{12} = 0$ $(j = 1, 2)$, so that

$$(17)\qquad (a_{21} : a_{11} - a_{22} : -a_{12}) = (1 : \xi_1 + \xi_2 : \xi_1\xi_2).$$

Now because $Q((a_{21}y + a_{22})^{-1}(a_{11}y + a_{12}))(a_{21}y + a_{22})^d$ is divisible by $Q(y)$ and obviously has the same degree as $Q(y)$, the quotient of these two polynomials is in $\mathfrak{F}$; since $Q(y) = (y - \xi_1)^{h_1}(y - \xi_2)^{h_2}$, an easy computation shows that this quotient is $(a_{11} - a_{21}\xi_1)^{h_1}(a_{11} - a_{21}\xi_2)^{h_2}$. But

$$(a_{11} - a_{21}\xi_1)(a_{11} - a_{21}\xi_2) = |\sigma|,$$

so that

$$Q((a_{21}y + a_{22})^{-1}(a_{11}y + a_{12}))(a_{21}y + a_{22})^d = |\sigma|^{h_1}(-a_{21}\xi_2 + a_{11})^{h_1 - h_2}Q(y).$$

It follows from (14) that

$$|\sigma|^{h_1}(-a_{21}\xi_2 + a_{11})^{h_2 - h_1}(|\sigma| P_i(y) + \Lambda_i(y)Q(y)) = P_i((a_{21}y + a_{22})^{-1}(a_{11}y + a_{12}))(a_{21}y + a_{22})^{d+2}.$$

Replacing y by ξ_j we find that

$$|\sigma|^{h_1+1}(-a_{21}\xi_2 + a_{11})^{h_2 - h_1}P_i(\xi_j) = P_i(\xi_j)(a_{21}\xi_j + a_{22})^{d+2}.$$

Now, for at least one value of i, $P_i(\xi_j) \neq 0$, for otherwise $P_1, \cdots, P_m, Q$ would have the common factor $y - \xi_j$. Therefore

$$|\sigma|^{h_1+1}(-a_{21}\xi_2 + a_{11})^{h_2 - h_1} = (a_{21}\xi_j + a_{22})^{d+2}.$$

Since the left member here is the same for both values of j, the same must be true for the right member, that is $(a_{21}\xi_1 + a_{22})^{d+2} = (a_{21}\xi_2 + a_{22})^{d+2}$, so that $a_{21}\xi_1 + a_{22} = \mu(a_{21}\xi_2 + a_{22})$, where μ is one of the $(d+2)$-th roots of unity. But this equation and (17) together admit only a finite number of solutions $(a_{11} : a_{12} : a_{21} : a_{22})$. This contradicts the infinite number of possibilities for σ, and proves that Q can not have two distinct zeros.

Suppose now that Q has precisely one zero ξ. If we set $\bar{\theta} = \theta - \xi$, so that $\mathfrak{G} = \mathfrak{F}(\bar{\theta})$, then $\delta_i\bar{\theta} = \bar{\theta}^{-d}(P_i(\bar{\theta} + \xi) - \delta_i\xi \cdot \bar{\theta}^d) = \bar{\theta}^{-d}\bar{P}_i(\bar{\theta})$, where $\bar{P}_i(y) = P_i(y + \xi) - \delta_i\xi \cdot y^d$; whenever $P_i \neq 0$ then $\bar{P}_i \neq 0$ and is of degree $d + 2$; if $P_i = 0$ then, choosing some j such that P_j is not divisible by $y - \xi$ and letting P_{ji} denote the polynomial obtained by replacing each coefficient ϕ in P_j by $\delta_i\phi$, we find that

$$\begin{aligned}
0 &= \delta_j(Q(\theta)^{-1}P_i(\theta)) = \delta_j\delta_i\theta = \delta_i\delta_j\theta \\
&= \delta_i(Q(\theta)^{-1}P_j(\theta)) = \delta_i((\theta - \xi)^{-d}P_j(\theta)) \\
&= -d(\theta - \xi)^{-d-1}((\theta - \xi)^{-d}P_i(\theta) - \delta_i\xi)P_j(\theta) \\
&\quad + (\theta - \xi)^{-d}(P_{ji}(\theta) + P'_j(\theta)(\theta - \xi)^{-d}P_i(\theta)) \\
&= d(\theta - \xi)^{-d-1}\delta_i\xi \cdot P_j(\theta) + (\theta - \xi)^{-d}P_{ji}(\theta),
\end{aligned}$$

so that $d\delta_i\xi \cdot P_j$ is divisible by $y - \xi$, whence $\delta_i\xi = 0$, and $\bar{P}_i = 0$. Therefore we lose no generality in assuming that $Q = y^d$. For every automorphism σ of $\mathfrak{G}$ over $\mathfrak{F}$ the transformation (16) then leaves 0 invariant, so that $a_{12} = 0$, and

$$Q((a_{21}y + a_{22})^{-1}(a_{11}y + a_{12}))(a_{21}y + a_{22})^d = {a_{11}}^d y^d = {a_{11}}^d Q(y).$$

It follows from (14) that

$$(18)\quad {a_{11}}^d(|\,\sigma\,|\,P_i(y) + A_i(y)y^d) = P_i((a_{21}y + a_{22}))^{-1}a_{11}y)(a_{21}y + a_{22})^{d+2}.$$

Replacing y by 0 here we find that

$$a_{11}{}^d\,|\,\sigma\,|\,P_i(0) = P_i(0){a_{22}}^{d+2}.$$

But $|\,\sigma\,| = a_{11}a_{22} \neq 0$, and $P_i(0) \neq 0$ for some i; therefore ${a_{11}}^{d+1} = {a_{22}}^{d+1}$, whence $a_{22} = \mu a_{11}$, where now μ is one of the $(d+1)$-th roots of unity. By (15) we thus find that

$$A_i(y) = (a_{21}\delta_i a_{11} - a_{11}\delta_i a_{21})y^2 = -\,{a_{11}}^2\delta_i({a_{11}}^{-1}a_{21}) \cdot y^2,$$

so that (18) becomes

$$\mu P_i(y) - \delta_i({a_{11}}^{-1}a_{21}) \cdot y^{d+2} = P_i(({a_{11}}^{-1}a_{21}y + \mu)^{-1}y)({a_{11}}^{-1}a_{21}y + \mu)^{d+2}.$$

Equating coefficients of y here we obtain

$$(d+2)P_i(0){a_{11}}^{-1}a_{21} = P'_i(0)(\mu - 1).$$

Since $P_i(0) \neq 0$ for some i, and since we already know that $a_{12} = 0$ and $a_{22} = \mu a_{11}$, this contradicts the fact that the number of automorphisms σ is infinite. Therefore Q does not have a zero, so that $Q = 1$, and each P_i which is different from 0 has degree 2.

Thus we may write

$$P_i = p_{i0} + p_{i1}y + p_{i2}y^2 \qquad (p_{ij} \in \mathfrak{F},\ p_{i2} \neq 0 \text{ if } P_i \neq 0).$$

From this an easy computation shows that

$$\delta_j\delta_i\theta = \delta_j p_{i0} + p_{j0}p_{i1} + (\delta_j p_{i1} + 2p_{j0}p_{i2} + p_{j1}p_{i1})\theta + (\delta_j p_{i2} + p_{j2}p_{i1} + 2p_{j1}p_{i2})\theta^2 + 2p_{j2}p_{i2}\theta^3.$$

Since $\delta_j\delta_i = \delta_i\delta_j$ this implies that

$$(19)\quad \begin{cases} \delta_j p_{i0} + p_{j0}p_{i1} = \delta_i p_{j0} + p_{i0}p_{j1}, \\ \delta_j p_{i1} + 2p_{j0}p_{i2} = \delta_i p_{j1} + 2p_{i0}p_{j2}, \\ \delta_j p_{i2} + p_{j1}p_{i2} = \delta_i p_{j2} + p_{i1}p_{j2}. \end{cases}$$

Let σ_0 be any automorphism of $\mathscr{G}$ over $\mathscr{F}$ such that $\sigma_0\theta \neq \theta$. It is easy to verify that the conditions

$$\delta_i(\tfrac{1}{2}p_{j1} + p_{j2}\theta) = \delta_j(\tfrac{1}{2}p_{i1} + p_{i2}\theta)$$

hold for all i and j. These conditions imply that the differential ideal $[\delta_1 z + (\frac{1}{2}p_{11} + p_{12}\theta)z, \cdots, \delta_m z + (\frac{1}{2}p_{m1} + p_{m2}\theta)z]$ in $\mathscr{G}\{z\}$ does not contain z and thus has a zero $\zeta_1 \neq 0$; it is not difficult to see that we may take ζ_1 so that the field of constants of $\mathscr{G}\langle\zeta_1\rangle$ is $\mathscr{C}$. Similarly, there exists a zero $\zeta_2 \neq 0$ of the differential ideal

$$[\delta_1 z + (\tfrac{1}{2}p_{11} + p_{12}\sigma_0\theta)z, \cdots, \delta_m z + (\tfrac{1}{2}p_{m1} + p_{m2}\sigma_0\theta)z]$$

in $\mathscr{G}\langle\zeta_1\rangle\{z\}$ such that the field of constants of $\mathscr{G}\langle\zeta_1, \zeta_2\rangle$ is $\mathscr{C}$. For any pair θ_1, θ_2 of operators of the form $\delta_1^{i_1} \cdots \delta_m^{i_m}$ we write

$$W_{\theta_1,\theta_2}(z_1, z_2) = \theta_1 z_1 \cdot \theta_2 z_2 - \theta_2 z_1 \cdot \theta_1 z_1.$$

It is easy to verify that

$$W_{1,\delta_i}(\zeta_1, \zeta_2) = p_{i2}\zeta_1\zeta_2(\theta - \sigma_0\theta), \tag{20}$$

which is different from 0 for at least one value of i, and that

$$W_{\theta_1,\theta_2}(\zeta_1, \zeta_2)\zeta_1^{-1}\zeta_2^{-1}(\theta - \sigma_0\theta)^{-1} \in \mathscr{F}$$

for every pair θ_1, θ_2 of operators of order ≤ 2. Therefore (Kolchin [6], § 3) $\mathscr{H} = \mathscr{F}\langle\zeta_1, \zeta_2\rangle$ is a Picard-Vessiot extension of $\mathscr{F}$. Of course $\mathscr{G} \subseteq \mathscr{H}$.

We denote the group of all automorphisms of $\mathscr{H}$ over $\mathscr{F}$ by $\mathfrak{H}$. By the Picard-Vessiot theory (Kolchin [6]) each element $\tau \in \mathfrak{H}$ may be identified with an element $(b_{ij}) = (b_{ij})_{1\leq i\leq 2,\ 1\leq j\leq 2}$ of an algebraic matric group over $\mathscr{C}$ by means of equations

$$\tau\zeta_j = b_{1j}\zeta_1 + b_{2j}\zeta_2 \qquad (j = 1, 2).$$

Let σ be any automorphism of $\mathscr{G}$ over $\mathscr{F}$ distinct from the identity and σ_0; σ can (Kolchin [6], § 3, Theorem 3) be extended to an element $\tau \in \mathfrak{H}$. Writing $\tau = (b_{ij})$ we have

$$\begin{aligned}
0 = \tau 0 &= \tau(\delta_i\zeta_1 + (\tfrac{1}{2}p_{i1} + p_{i2}\theta)\zeta_1) \\
&= b_{11}\delta_i\zeta_1 + b_{21}\delta_i\zeta_2 + (\tfrac{1}{2}p_{i1} + p_{i2}\sigma\theta)(b_{11}\zeta_1 + b_{21}\zeta_2) \\
&= -b_{11}(\tfrac{1}{2}p_{i1} + p_{i2}\theta)\zeta_1 - b_{21}(\tfrac{1}{2}p_{i1} + p_{i2}\sigma_0\theta)\zeta_2 \\
&\qquad + (\tfrac{1}{2}p_{i1} + p_{i2}\sigma\theta)(b_{11}\zeta_1 + b_{21}\zeta_2) \\
&= b_{11}p_{i2}(\sigma\theta - \theta)\zeta_1 + b_{21}p_{i2}(\sigma\theta - \sigma_0\theta)\zeta_2;
\end{aligned}$$

since $\det(b_{ij}) \neq 0$ this implies that $b_{11}b_{21} \neq 0$ and $\zeta_1^{-1}\zeta_2 \in \mathscr{G}$.

A straightforward computation shows that $\delta_i(\zeta_1\zeta_2(\theta - \sigma_0\theta)) = 0$ $(1 \leq i \leq m)$, so that $\zeta_1\zeta_2(\theta - \sigma_0\theta) = c \in \mathscr{C}$. From (20) we see at once that $W_{1,\delta_i}(\zeta_1, \zeta_2) = cp_{i2} \in \mathfrak{F}$, so that

$$\det(b_{ij}) \cdot W_{1,\delta_i}(\zeta_1, \zeta_2) = W_{1,\delta_i}(\tau\zeta_1, \tau\zeta_2) = \tau W_{1,\delta_i}(\zeta_1, \zeta_2) = W_{1,\delta_i}(\zeta_1, \zeta_2)$$

whence $\det(b_{ij}) = 1$ for all $\tau \in \mathfrak{H}$. Also, $\zeta_1^{-2} W_{1,\delta_i}(\zeta_1, \zeta_2) = \delta_i(\zeta_1^{-1}\zeta_2) \in \mathfrak{G}$, so that $\zeta_1^2 \in \mathfrak{G}$, whence $\zeta_1^2, \zeta_1\zeta_2, \zeta_2^2 \in \mathfrak{G}$. Since $\delta_i\zeta_1^2 = -2\zeta_1^2(\frac{1}{2}p_{i1} + p_{i2}\theta)$, it follows that $\theta \in \mathfrak{F}\langle\zeta_1^2, \zeta_1\zeta_2, \zeta_2^2\rangle$, so that $\mathfrak{G} = \mathfrak{F}\langle\zeta_1^2, \zeta_1\zeta_2, \zeta_2^2\rangle$. Therefore $\partial^0\mathfrak{H}/\mathfrak{F} = \partial^0\mathfrak{G}/\mathfrak{F} = 1$, whence $\dim \mathfrak{H} = 1$.

It follows that the component of the identity $\mathfrak{H}^0$ is reducible either to diagonal form or else to special triangular form, that is, there exist two linear combinations ω_1, ω_2 of ζ_1, ζ_2 over $\mathscr{C}$, which are linearly independent over constants, such that either for every $\tau \in \mathfrak{H}^0$ there exists a nonzero $b \in \mathscr{C}$ for which $\tau\omega_1 = b\omega_1$, $\tau\omega_2 = b^{-1}\omega_2$, or else for every $\tau \in \mathfrak{H}^0$ there exists a $b \in \mathscr{C}$ for which $\tau\omega_1 = \omega_1$, $\tau\omega_2 = b\omega_1 + \omega_2$.

In the former case $\tau(\omega_j^{-1}\delta_i\omega_j) = \omega_j^{-1}\delta_i\omega_j$ for every $\tau \in \mathfrak{H}^0$ so that $\omega_j^{-1}\delta_i\omega_j$ is algebraic over $\mathfrak{F}$; but

$$\omega_j^{-1}\delta_i\omega_j = \tfrac{1}{2}\omega_j^{-2}\delta_i(\omega_j^2) \in \mathfrak{F}\langle\zeta_1^2, \zeta_1\zeta_2, \zeta_2^2\rangle = \mathfrak{G}$$

and $\mathfrak{F}$ is relatively algebraically closed in $\mathfrak{G}$, so that $\omega_j^{-1}\delta_i\omega_j \in \mathfrak{F}$, that is, ω_j is exponential over $\mathfrak{F}$. Therefore for every τ in $\mathfrak{H}$ there exists a nonzero $b \in \mathscr{C}$ such that $\tau\omega_1 = b\omega_1$, $\tau\omega_2 = b^{-1}\omega_2$. Consequently $\omega_1\omega_2$ is invariant under every $\tau \in \mathfrak{H}$ and belongs to $\mathfrak{F}$, so that

$$\mathfrak{G} = \mathfrak{F}\langle\zeta_1^2, \zeta_1\zeta_2, \zeta_2^2\rangle = \mathfrak{F}\langle\omega_1^2, \omega_1\omega_2, \omega_2^2\rangle = \mathfrak{F}\langle\omega_1^2\rangle.$$

Setting $\alpha = \omega_1^2$ we see that $\alpha^{-1}\delta_i\alpha = 2\omega_1^{-1}\delta_i\omega_1 \in \mathfrak{F}$, so that α is exponential over $\mathfrak{F}$, and also that $\mathfrak{G} = \mathfrak{F}\langle\alpha\rangle$.

In the latter case $\tau\omega_1 = \omega_1$ for every $\tau \in \mathfrak{H}^0$, so that ω_1 is algebraic over $\mathfrak{F}$ whence, since $\omega_1^2 \in \mathfrak{F}\langle\zeta_1^2, \zeta_1\zeta_2, \zeta_2^2\rangle = \mathfrak{G}$, we have $\omega_1^2 \in \mathfrak{F}$. Since $W_{1,\delta_i}(\zeta_1, \zeta_2) = cp_{i2} \in \mathfrak{F}$, we also have $W_{1,\delta_i}(\omega_1, \omega_2) \in \mathfrak{F}$, so that $\delta_i(\omega_1^{-1}\omega_2) = \omega_1^{-2} W_{1,\delta_i}(\omega_1, \omega_2) \in \mathfrak{F}$. Therefore if we set $\alpha = \omega_1^{-1}\omega_2$ then α is primitive over $\mathfrak{F}$ and $\mathfrak{G} = \mathfrak{F}\langle\omega_1^2, \omega_1\omega_2, \omega_2^2\rangle = \mathfrak{F}\langle\omega_1^2, \omega_1^2\alpha, \omega_1^2\alpha^2\rangle = \mathfrak{F}\langle\alpha\rangle$.

This completes the treatment of the case of genus 0.

11. The proof concluded. Case of genus 1. We consider now the remaining case in which $\mathfrak{F}$ is algebraically closed and $\mathfrak{G}$ is of genus 1. It is known (for example see Chevalley [2], Chapter II, § 3) that in this case there exist two elements α, β in $\mathfrak{G}$ such that $\mathfrak{G} = \mathfrak{F}(\alpha, \beta)$ and $\beta^2 = P(\alpha)$,

where P is a cubic polynomial in $\mathfrak{F}[y]$ which does not have a multiple root. Replacing α, β by suitable elements $a\alpha + b$, $c\beta$ ($a, b, c \in \mathfrak{F}$), we lose no generality in supposing that

$$\beta^2 = 4\alpha^3 - g_2\alpha - g_3, \tag{21}$$

where $g_2 \in \mathfrak{C}$, $g_3 \in \mathfrak{F}$, and $27g_3^2 - g_2^3 \neq 0$.

We shall prove that then $g_3 \in \mathfrak{C}$, $\mathfrak{G} = \mathfrak{F}\langle\alpha\rangle$, and there exist elements $a_1, \cdots, a_m \in \mathfrak{F}$ (not all 0) such that

$$(\delta_i\alpha)^2 = a_i^2(4\alpha - g_2\alpha - g_3), \qquad (1 \leq i \leq m \tag{22}$$

(so that α is weierstrassian over $\mathfrak{F}$). This will complete the proof of Theorem 7.

We begin by observing that α is transcendental over $\mathfrak{F}$; since the field of constants of $\mathfrak{G}$ is $\mathfrak{C}$, α is not a constant. If $\delta_i\alpha = 0$ then (22) holds with $a_i = 0$. Let i be any index such that $\delta_i\alpha \neq 0$; in what follows we shall keep i fixed, and for every element ξ of $\mathfrak{G}$ we shall denote $\delta_i\xi$ by ξ'.

Clearly there exist polynomials $A, B, C \in \mathfrak{F}[y]$, without common divisor and with the leading coefficient in C equal to 1, such that

$$\alpha' = \frac{A(\alpha) + B(\alpha)\beta}{C(\alpha)}. \tag{23}$$

Applying δ_i to both members of (21) we find that $2\beta\beta' = (12\alpha^2 - g_2)\alpha' - g'_3$; from this, (21), and (23) we obtain

$$\beta' = \frac{(12\alpha^2 - g_2)(4\alpha^3 - g_2\alpha - g_3)B(\alpha) + ((12\alpha^2 - g_2)A(\alpha) - g'_3C(\alpha))\beta}{2C(\alpha)(4\alpha^3 - g_2\alpha - g_3)}. \tag{24}$$

Now $(1:\alpha:\beta)$ is a point of the group variety $\mathfrak{W}(g_2, g_3)$ defined in § 6. Since there exist infinitely many automorphisms over $\mathfrak{F}$ of the differential field $\mathfrak{G}$, there exist infinitely many such automorphisms σ such that

$$(1:\sigma\alpha:\sigma\beta) = (1:a:b)(1:\alpha:\beta), \tag{25}$$

where $a, b \in \mathfrak{F}$ and $(1:a:b) \in \mathfrak{W}(g_2, g_3)$,[10] that is, such that

---

[10] This follows from the known fact that, in the group of all automorphisms of the algebraic function field $\mathfrak{G}$, the subgroup of those automorphisms for which an equation of the form (25) holds is of finite index. To see this observe that this subgroup acts transitively on the set of all places of the function field $\mathfrak{G}$, and that there exists a place $\mathfrak{p}$ at which α has a pole of order 2 and β has a pole of order 3; the above fact is then a consequence of a second fact, namely that the group of all automorphisms of the function field $\mathfrak{G}$ which leave $\mathfrak{p}$ invariant is finite. This second fact is in turn an easy

$$(26)\qquad \begin{cases} \sigma\alpha = -\alpha - a + \frac{1}{4}\left(\dfrac{\beta - b}{\alpha - a}\right)^2, \\ \sigma\beta = -\frac{1}{2}(\beta + b) + \frac{3}{2}\dfrac{\beta - b}{\alpha - a} - \frac{1}{4}\left(\dfrac{\beta - b}{\alpha - a}\right)^3. \end{cases}$$

These equations may, with the help of (21), be rewritten in the form

$$(27)\qquad \begin{cases} \sigma\alpha = \dfrac{4a\alpha^2 + (4a^2 - g_2)\alpha - (g_2 a + 2g_3) - 2b\beta}{4(\alpha - a)^2}, \\ \sigma\beta = \dfrac{(4\alpha^3 + 12a\alpha^2 - 3g_2\alpha - g_2 a - 4g_3)b + ((-12a^2 + g_2)\alpha - 4a^3 + 3g_2 a + 4g_3)\beta}{4(\alpha - a)^3}. \end{cases}$$

Applying δ_i to the first equation (26), and making use of (21), (23), and (24), we obtain

$$(\sigma\alpha)' = (U + V\beta)4^{-1}(\alpha - a)^{-1}(4\alpha^3 - g_2\alpha - g_3)^{-1}C(\alpha)^{-1},$$

where

$$\begin{aligned} (28)\quad U = {} & -4a'(\alpha - a)^3(4\alpha^3 - g_2\alpha - g_3)C(\alpha) - 4(\alpha - a)^3(4\alpha^3 - g_2\alpha - g_3)A(\alpha) \\ & + (\alpha - a)(12\alpha^2 - g_2)(4\alpha^3 - g_2\alpha - g_3)A(\alpha) - g'_3(\alpha - a)(4\alpha^3 - g_2\alpha - g_3)C(\alpha) \\ & - b(\alpha - a)(12\alpha^2 - g_2)(4\alpha^3 - g_2\alpha - g_3)B(\alpha) \\ & + ((12a^2 - g_2)a' - g'_3)(\alpha - a)(4\alpha^3 - g_2\alpha - g_3)C(\alpha) \\ & + 2b(4\alpha^3 - g_2\alpha - g_3)(A(\alpha) - a'C(\alpha)) - 2(4\alpha^3 - g_2\alpha - g_3)^2B(\alpha), \end{aligned}$$

and

$$\begin{aligned} (29)\quad V = {} & -4(\alpha - a)^3(4\alpha^3 - g_2\alpha - g_3)B(\alpha) + (\alpha - a)(12\alpha^2 - g_2)(4\alpha^3 - g_2\alpha - g_3)B(\alpha) \\ & - 2b'(\alpha - a)(4\alpha^3 - g_2\alpha - g_3)C(\alpha) - b(\alpha - a)((12\alpha^2 - g_2)A(\alpha) - g'_3C(\alpha)) \\ & - 2(4\alpha^3 - g_2\alpha - g_3)(A(\alpha) - a'C(\alpha)) + 2b(4\alpha^3 - g_2\alpha - g_3)B(\alpha). \end{aligned}$$

On the other hand, by (23)

$$\sigma(\alpha') = \frac{A(\sigma\alpha) + B(\sigma\alpha)\sigma\beta}{C(\sigma\alpha)}.$$

Since $(\sigma\alpha)' = \sigma(\alpha')$ we therefore find, with the help of (27), that

---

consequence of the Riemann-Roch theorem; indeed if σ_0 is any automorphism which leaves $\mathfrak{p}$ invariant then $\sigma_0\alpha$ has a pole of order 2 at $\mathfrak{p}$ and $\sigma_0\beta$ has a pole of order 3 there, whence (for example see Chevalley [2], chapter II, corollary to theorem 6) $\sigma_0\alpha = c_1\alpha + c_2$, $\sigma_0\beta = c_3\beta + c_4\alpha + c_5$, where $c_1, \cdots, c_5 \in \mathfrak{F}$ and $c_1c_3 \neq 0$; since σ_0 must preserve equation (21), an easy computation shows that $c_2 = c_4 = c_5 = 0$, $c_1^3 - c_3^2 = 0$, $g_2(c_1 - c_3^2) = 0$, $g_3(c_3^2 - 1) = 0$, so that there are only a finite number of possibilities for σ_0. For this short proof I am indebted to M. Rosenlicht.

$$(30)\qquad 4(\alpha-a)^3(4\alpha^3-g_2\alpha-g_3)C(\alpha)\Big[A\Big(W-\frac{b\beta}{2(\alpha-a)^2}\Big)$$
$$+B\Big(W-\frac{b\beta}{2(\alpha-a)^2}\Big)$$
$$\times\frac{(4\alpha^3+12a\alpha^2-3g_2\alpha-g_2a-4g_3)b+((-12a^2+g_2)\alpha-4a^3+3g_2a+4g_3)\beta}{4(\alpha-a)^3}\Big]$$
$$=C\Big(W-\frac{b\beta}{2(\alpha-a)^2}\Big)(U+V\beta),$$

where

$$W=\frac{4a\alpha^2+(4a^2-g_2)\alpha-(g_2a+2g_3)}{4(\alpha-a)^2}.$$

Because of (21) the left hand member here is a linear combination of 1 and β with coefficients which belong to $\mathfrak{F}(\alpha)$, have denominators that are powers of $\alpha-a$, and have numerators that are divisible by $(4\alpha^3-g_2\alpha-g_3)C(\alpha)$. The right hand member can also be expressed as a linear combination of 1 and β, the coefficient of 1 being

$$(31)\qquad \sum_{j\geq 0}\frac{1}{(2j)!}C^{(2j)}(W)\,\frac{(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^j}{2^{2j}(\alpha-a)^{4j}}U$$
$$-\sum_{j\geq 0}\frac{1}{(2j+1)!}C^{(2j+1)}(W)\,\frac{(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^{j+1}}{2^{2j+1}(\alpha-a)^{4j+2}}bV,$$

and the coefficient of β being

$$(32)\qquad \sum_{j\geq 0}\frac{1}{(2j)!}C^{(2j)}(W)\,\frac{(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^j}{2^{2j}(\alpha-a)^{4j}}V$$
$$-\sum_{j\geq 0}\frac{1}{(2j+1)!}C^{(2j+1)}(W)\,\frac{(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^j}{2^{2j+1}(\alpha-a)^{4j+2}}bU.$$

Therefore (31) and (32) are both expressible as quotients in which the denominator is a power of $\alpha-a$ and the numerator is a polynomial in $\mathfrak{F}[\alpha]$ divisible by $(4\alpha^3-g_2\alpha-g_3)C(\alpha)$.

Observing from (28) that U is divisible by $4\alpha^3-g_2\alpha-g_3$, and from (29) that each term of V is so divisible except for

$$-b(\alpha-a)((12\alpha^2-g_2)A(\alpha)-g'_3C(\alpha)),$$

and recalling that α is transcendental over $\mathfrak{F}$, we conclude, on substituting for α in (32) any root e of $4y^3-g_2y-g_3$, that

$$C\Big(\frac{4ae^2+(4a^2-g_2)e-(g_2a+2g_3)}{4(e-a)^2}\Big)$$
$$\times b(e-a)((12e^2-g_2)A(e)-g'_3C(e))=0.$$

Since this is true for infinitely many points $(1:a:b)$ of the curve $\boldsymbol{W}(g_2, g_3)$, this implies that $(12e^2 - g_2)A(e) - g'_3 C(e) = 0$. Because this equation holds for each of the three roots e of $4y^3 - g_2 y - g_3$ we conclude that

$$(33) \qquad (12y^2 - g_2)A(y) - g'_3 C(y) \equiv 0 \pmod{4y^3 - g_2 y - g_3}.$$

Returning now to (31) we see that if r denotes the degree of $C(y)$ and if we multiply (31) by $(\alpha - a)^{2r}$ then we obtain a polynomial in $\mathfrak{F}[\alpha]$ divisible by $C(\alpha)$, that is, we have a congruence in $\mathfrak{F}[\alpha]$ of the form

$$LU + MbV \equiv 0 \pmod{C(\alpha)}.$$

Since every subgroup of $\boldsymbol{W}(g_2, g_3)$ which contains $(1:a:b)$ also contains $1:a:-b) = (1:a:b)^{-1}$, this congruence continues to hold if we replace b by $-b$. Using the two congruences (one with b and one with $-b$), and observing from (28) and (29) that

$$\begin{aligned} U \equiv{}& (-4(\alpha-a)^3 + (\alpha-a)(12\alpha^2 - g_2) + 2b)(4\alpha^3 - 2g_2\alpha - g_3)A(\alpha) \\ &- (b(\alpha-a)(12\alpha^2 - g_2) + 2(4\alpha^3 - g_2\alpha - g_3))(4\alpha^3 - g_2\alpha - g_3)B(\alpha) \\ &\pmod{C(\alpha)} \end{aligned}$$

and

$$\begin{aligned} V \equiv{}& -(b(\alpha-a)(12\alpha^2 - g_2) + 2(4\alpha^3 - g_2\alpha - g_3))A(\alpha) \\ &+ (-4(\alpha-a)^3 + (\alpha-a)(12\alpha^2 - g_2) + 2b)(4\alpha^3 - g_2\alpha - g_3)B(\alpha) \\ &\pmod{C(\alpha)}, \end{aligned}$$

we obtain the following two congruences in $\mathfrak{F}[\alpha]$, in which b no longer appears:

$$\begin{aligned} (34) \quad & \Big[\sum_{j\geqq 0} \frac{1}{(2j)!2^{2j}} C^{(2j)}(W)(\alpha-a)^{2r-4j} \\ &\quad \times (4a^3 - g_2 a - g_3)^j (4\alpha^3 - g_2\alpha - g_3)^j (12\alpha^2 - g_2 - 4(\alpha-a)^2) \\ &\quad + \sum_{j\geqq 0} \frac{1}{(2j+1)!2^{2j+1}} C^{(2j+1)}(W)(\alpha-a)^{2r-4j-2} \\ &\quad \times (4a^3 - g_2 a - g_3)^{j+1} (4\alpha^3 - g_2\alpha - g_3)^j (12\alpha^2 - g_2)\Big](\alpha-a)A(\alpha) \\ &\quad - \Big[\sum_{j\geqq 0} \frac{1}{(2j)!2^{2j}} C^{(2j)}(W)(\alpha-a)^{2r-4j} \\ &\quad \times (4a^3 - g_2 a - g_3)^j (4\alpha^3 - g_2\alpha - g_3)^{j+1} \\ &\quad + \sum_{j\geqq 0} \frac{1}{(2j+1)!2^{2j+1}} C^{(2j+1)}(W)(\alpha-a)^{2r-4j-2} \\ &\quad \times (4a^3 - g_2 a - g_3)^{j+1} (4\alpha^3 - g_2\alpha - g_3)^{j+1}\Big] 2B(\alpha) \\ &\qquad \equiv 0 \pmod{C(\alpha)}, \end{aligned}$$

and

$$
\begin{aligned}
(35)\quad & \Big[\sum_{j\geqq 0}\frac{1}{(2j)\,!2^{2j}}C^{(2j)}(W)\,(\alpha-a)^{2r-4j} \\
& \quad\times(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^j \\
& \quad+\sum_{j\geqq 0}\frac{1}{(2j+1)\,!2^{2j+1}}C^{(2j+1)}(W)\,(\alpha-a)^{2r-4j-2} \\
& \quad\times(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^{j+1}\Big]2A(\alpha) \\
& -\Big[\sum_{j\geqq 0}\frac{1}{(2j)\,!2^{2j}}C^{(2j)}(W)\,(\alpha-a)^{2r-4j} \\
& \quad\times(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^j(12\alpha^2-g_2) \\
& \quad+\sum_{j\geqq 0}\frac{1}{(2j+1)\,!2^{2j+1}}C^{(2j+1)}(W)\,(\alpha-a)^{2r-4j-2} \\
& \times(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^{j+1}(12\alpha^2-g_2-4(\alpha-a)^2)\Big](\alpha-a)B(\alpha) \\
& \qquad\equiv 0\ (\mathrm{mod}\ C(\alpha)).
\end{aligned}
$$

Since the congruences (34) and (35) hold for infinitely many points $(1:a:b)\,\varepsilon\,\boldsymbol{W}(g_2,g_3)$ with $a,b\,\varepsilon\,\boldsymbol{\mathfrak{F}}$, that is for infinitely many elements $a\,\varepsilon\,\boldsymbol{\mathfrak{F}}$, they must hold for all elements a. In particular, they hold for a equal to α. Now the leading coefficient in $C(y)$ is 1; therefore, when we replace a by α, $C^{(k)}(W)(\alpha-a)^{2r-2k}$ becomes $(r\,!(r-k)\,!)2^{r-k}(4\alpha^3-g_2\alpha-g_3)^{r-k}$. Consequently, when a is replaced by α, (34) becomes

$$
\begin{aligned}
& -\Big[\sum_{j\geqq 0}\binom{r}{2j}2^{r-4j}(4\alpha^2-g_2\alpha-g_3)^{r+1} \\
& \quad+\sum_{j\geqq 0}\binom{r}{2j+1}2^{r-4j-2}(4\alpha^3-g_2\alpha-g_3)^{r+1}\Big]2B(\alpha)\equiv 0 \pmod{C(\alpha)},
\end{aligned}
$$

so that

$$B(\alpha)\,(4\alpha^3-g_2\alpha-g_3)^{r+1}\equiv 0 \pmod{C(\alpha)}.$$

In the same way, on replacing a by α in (35), we find that

$$A(\alpha)\,(4\alpha^3-g_2\alpha-g_3)^{r}\equiv 0 \pmod{C(\alpha)}.$$

Since A, B, C have no common divisor, it follows that

$$(4y^3-g_2y-g_3)^{r+1}\equiv 0 \pmod{C(y)}.$$

Therefore if C were of degree >0 then C would have a root e in common with $4y^3-g_2y-g_3$, and we would obtain, on replacing α by e in (35)

$$C\left(\frac{4ae^2+4(a^2-g_2)e-(g_2a+2g_3)}{4(e-a)^2}\right)(e-a)^{2r}$$
$$\times\,(2A(e)-(e-a)(12e^2-g_2)B(e))=0,$$

so that we would have $A(e)=0$, $B(e)=0$, contradicting the fact that A, B, C are without common divisor. Consequently

$$C(y)=1. \tag{36}$$

Now A and B are not both 0, for otherwise by (23) we would have $\delta_i\alpha=\alpha'=0$, contrary to assumption. Let p denote the maximum of the degrees of A and B, and denote the coefficients of y^p in A and B by c_A and c_B respectively. Suppose p were >0. Multiplying both sides of (30) by $(\alpha-a)^{2p}$, then expressing each side as a linear combination, over $\mathfrak{F}(\alpha)$, of 1 and β, and then equating coefficients of 1, we would obtain an equation in $\mathfrak{F}[\alpha]$; if in this equation we equated the coefficients, right and left of α^{3p+6} we would obtain

$$32c_A-(48b+32)c_B=0;$$

since this would hold for infinitely many points $(1:a:b)\,\varepsilon\,\mathfrak{W}(g_2,g_3)$, that is, for infinitely many values of b, we would have $c_A=c_B=0$, which is impossible. Therefore $p=0$, so that $A=c_A\,\varepsilon\,\mathfrak{F}$, $B=c_B\,\varepsilon\,\mathfrak{F}$. By (36) and (33) we further conclude that $c_A=0$ and $g'_3=0$, whence from (23) $\alpha'=c_B\beta$. Writing $c_B=a_i$ we see, by (21), that (22) holds whence $\mathfrak{y}=\mathfrak{F}(\alpha,\beta)=\mathfrak{F}\langle\alpha\rangle$. To complete the proof of the theorem it remains to show that $g_3\,\varepsilon\,\mathfrak{k}$. We have just seen that $\delta_ig_3=0$ for those values of i for which $\delta_i\alpha\neq 0$; we must prove that $\delta_kg_3=0$ for all values of k such that $\delta_k\alpha=0$. For such i and k we have

$$\delta_k((\delta_i\alpha)^2)=2\delta_i\cdot\alpha\cdot\delta_k\delta_i\alpha=2\delta_i\alpha\cdot\delta_i\delta_k\alpha=0$$

so that, because of (22),

$$0=\delta_k(a_i^2(4\alpha^3-g_2\alpha-g_3))=2a_i\cdot\delta_ka_i\cdot(4\alpha^3-g_2\alpha-g_3)+a_i^2(-\delta_kg_3);$$

since α is transcendental over $\mathfrak{F}$, this implies that $\delta_ka_i=0$ and $\delta_kg_3=0$. Thus $g_3\,\varepsilon\,\mathfrak{k}$, and the proof of Theorem 7 is complete.

COLUMBIA UNIVERSITY.

REFERENCES.

[1] N. Bourbaki, "*Algèbre,*" Chapters IV-V (Actualités scientifiques et industrielles 1102), Hermann et Cie., Paris, 1950.

[2] C. Chevalley, "*Introduction to the theory of algebraic functions of one variable,*" (Mathematical Surveys VI), American Mathematical Society, New York, 1951.

[3] E. R. Kolchin, "Algebraic matric groups and the Picard-Vessiot theory of homogeneous linear ordinary differential equations," *Annals of Mathematics*, vol. 49 (1948), pp. 1-42.

[4] ———, "Existence theorems connected with the Picard-Vessiot theory of homogeneous linear ordinary differential equations," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 927-932.

[5] ———, "On certain concepts in the theory of algebraic matric groups," *Annals of Mathematics*, vol. 49 (1948), pp. 774-789.

[6] ———, "Picard-Vessiot theory of partial differential fields," *Proceedings of the American Mathematical Society*, vol. 3 (1952), pp. 596-603.

[7] K. Iwasawa and T. Tamagawa, "On the group of automorphisms of a function field," *Journal of the Mathematical Society of Japan*, vol. 3 (1951), pp. 137-147.

[8] J. F. Ritt, "*Differential algebra,*" (American Mathematical Society Colloquium Publications, vol. 33), New York, 1950.

[9] A. Weil, "*Foundations of algebraic geometry,*" (ibid., vol. 29), New York, 1946.

[10] ———, "*Variétés abeliennes et courbes algébriques* (Actualités scientifiques et industrielles 1064), Hermann et Cie., Paris, 1948.

NETWORKS SATISFYING MINIMALITY CONDITIONS.* [1,2]

By R. Duncan Luce.

1. Introduction. A *network* N is a system of two finite sets M and $P \subset M \times M$, in which the elements $a, b, \cdots \varepsilon M$ are called the *nodes* and the elements $(ab), (ca), \cdots \varepsilon P$ are called the *links* of N. The number of nodes is denoted by m and the number of links by $p(N)$. If (ab) is a link of N, a is called the *initial node* and b the *end node* of the link (ab). Thus, a network is a binary relation over a finite set and is also a finite oriented graph in which there is at most one oriented arc from one node to another. Our viewpoint is primarily that of graph theory rather than algebra.

Let I be the set of all links of the form $(aa), a \varepsilon M$. If $P \cap I = 0$, N is called *non-reflexive*. We shall, without further mention, take the word network to mean non-reflexive network.

A *subnetwork* N' of a network N, denoted $N' \subset N$, is any network with $M' \subset M$ and $P' \subset P$. If $M' = M$, we say N' is a complete subnetwork of N. If $N' \subset N$, $N - N'$ is the network with nodes M and links $P - P'$ and it is said to be formed from N by the removal of the links P'. If N' has but one link (ab) we write $N - N' = N - (ab)$. Similarly, the network with nodes $M - M'$ and links $P \cap [(M - M') \times (M - M')]$ is said to be formed from N by the removal of the nodes M' (and the incident links). N' is a *supernetwork* of N if N is a complete subnetwork of N'. We shall write in this case $N' = N + (N' - N)$ and say that N' is formed from N by adding the links $P' - P$ to N. If $N' - N$ contains but one link (ab), we write $N' = N + (ab)$.

A pair of links (ab) and (ba) is called an *arc* ab, and any network composed entirely of arcs is isomorphic to a graph and so is called a graph.

A *q-chain* from a to b, denoted (ab, q), is an ordered sequence of q links $(ac_1), (c_1c_2), \cdots, (c_qb)$ in which no node is repeated, except possibly $a = b$. In the latter case the chain is called an (oriented) *circuit*. A network is

* Received April 6, 1953.

[1] This work has been supported in part by the Signal Corps, the Air Materiel Command, and the Office of Naval Research, U. S. A.

[2] The author is indebted to Josiah Macy, Jr., for many helpful discussions about this work and for a critical reading of the manuscript.

connected if there is a chain from every node to every other node; otherwise it is *disconnected.* Maximal connected subnetworks are called *components.*

In a previous paper [4], which will be referred to as (A), the following definitions were introduced:

A network has *degree* 0 if it is not connected; it has *degree* k, $k > 0$, if there exists $N' \subset N$ such that $p(N') = k$ and $N - N'$ is disconnected, but $N - N''$ is connected for all $N'' \subset N$ such that $p(N'') < k$.

A network N is *k-minimal* if the degree of $N - (ab)$ is $k - 1$ for every $(ab) \in N$. If N is 1-minimal and connected, it is called *minimal.*

In this paper we are concerned with three independent results which are each related to k-minimality. The definition is extended in a natural way to disconnected networks in Section 2 and these networks are completely characterized by Theorem 1. It is worth mentioning that the characterization problem for connected networks appears to be far more difficult. (The principal result of (A) is the solution to that problem for $k = 1$). In Section 3, the principal result is Theorem 4 which states that in a network of degree k, there is a set of at least k chains from any node to any other node, no two of which have a common link. This result is a generalization of a close analogue to the well known theorem of Menger that between any two nodes of a graph without a cut-node there are at least two chains that have no intermediate nodes in common. In the final section we turn to a generalization of transitivity. Connectedness and transitivity are each such strong requirements that combined they single out but one network—the case $P = M \times M$—so, in the presence of connectedness, transitivity must be weakened to be of interest. We require that every chain exceeding h links is "short-circuited" by a link, and that no chain of h or fewer links is short-circuited. It is shown that these connected networks fall into three classes: one having but one member which is of degree 2, the set of minimal networks, and a set non-minimal networks of degree 1 whose connected subnetworks also have degree 1.

2. $(-k)$-minimal networks.[3] To extend the above definitions of degree and minimality to disconnected networks, we simply interchange the roles of connected and disconnected as follows:

A network N has *degree* $(-k)$, $k \geqq 0$, if there exists a connected supernetwork N' of N such that $p(N' - N) = k + 1$, but every supernetwork N'' such that $p(N'' - N) < k + 1$ is disconnected.

---

[3] The author is indebted to Anatol Holt who suggested this problem to him.

A network N is (-0)-*minimal* if N is disconnected and for every $(ab) \notin N, a, b \in N$, $N + (ab)$ is connected; it is $(-k)$-minimal, $k \geqq 1$, if for every $(ab) \notin N, a, b \in N$, $N + (ab)$ has degree $(-k+1)$.

Following Dirac's terminology ([1], p. 347), we shall call a network with every possible link present a *complete graph*, and any component which is a complete graph is simply called complete. A node which is neither an end node nor an initial node of any link is called an *isolated node*.

LEMMA 1. *If N is a $(-k)$-minimal network with $m \geqq 3$ and $k \geqq 2$ and a is an isolated node of N, then $N' = N - a$ is either $(-k+1)$-minimal or a complete graph.*

Proof. If N' is not a complete graph, then since $m \geqq 3$, there exist $b, c \in N'$ such that $(bc) \notin N'$. For any such b and c, consider $N^* = N' + (bc)$. Let $-q$ be the degree of N^*, then the lemma is proved if we show $q = k - 2$. Let U be any set of k links which connects $N + (bc)$, and observe, since a is isolated, there exist $e, f \in N'$ such that $(ea), (af) \in U$. Now, $U + (ef) - (ea) - (af)$ connects N^*, so $q \leqq k - 2$. Suppose $q < k - 2$ and let U' be a set of $q + 1$ links which connects N^*. U' is non-empty, for otherwise N^* is connected, whence $N + (ba)$ is connected by adding (ac), and this implies N *is* (-1)-minimal, which contradicts $k \geqq 2$. Let $(e'f') \in U'$, then $U' - (e'f') + (e'a) + (af')$ connects $N + (bc)$ using only $(q+1) + 1 < k$ links, which is a contradiction.

THEOREM 1. *A network N is $(-k)$-minimal if and only if either*

(i) *N is a graph which consists of $k+1$ complete components having no link between any pair, or*

(ii) *N consists of a set X of nodes which form $k+1$ complete components having no link between any pair and a complete component Y such that either*

1. *$(xy) \in N$ and $(yx) \notin N$ for all $x \in X$ and $y \in Y$,*

or

2. *$(yx) \in N$ and $(xy) \notin N$ for all $x \in X$ and $y \in Y$.*

Proof. The sufficiency is obvious.

The condition is clearly necessary for $k = 0$, so we restrict the proof to $k \geqq 1$. Let N' be any component of N. If N' is an isolated node, it is complete. If N' has more than one node, we show it is complete: If there

exist $a, b \varepsilon N'$ such that $(ab) \notin N'$, and if U is any set of k links which connects $N + (ab)$, then for any $(cd) \varepsilon U$, $U - (cd)$ connects $N + (cd)$, since N' is already connected. This contradicts the assumption that N is $(-k)$-minimal.

If N', N'' are two components of N, we show that if $a \varepsilon N'$, $b \varepsilon N''$, and $(ab) \varepsilon N$, then $(a'b') \varepsilon N$ for any $a' \varepsilon N'$, $b' \varepsilon N''$: Suppose $(a'b') \notin N$, and let U be any set of k links connecting $N + (a'b')$. U connects N since N' and N'' are complete and $(ab) \varepsilon N$, which contradicts the assumption that N is $(-k)$-minimal.

Since the components of N are complete and since if there is one link from N' to N'' there are all possible links, it is sufficient to prove the theorem for networks having no components with more than one node.

First, $m \geqq k + 1$, for if not N can be connected as a circuit on all nodes with fewer than $k + 1$ links. If $m = k + 1$, no links are present, for if there were N could again be connected as a circuit using no more than $m - 1 = k$ links. In this case, N satisfies part (i) of the statement.

For networks with $m \geqq k + 2$, an induction on m will be used to show part (ii) holds. For $m = 3$, it is clear this is the case. Suppose $m > 3$ and part (ii) holds for $m' = m - 1$. If N has an isolated node a, then by Lemma 1, $N - a$ is $(-k + 1)$-minimal, so by the induction hypothesis (ii) holds for $N - a$, since (i) cannot. Thus, there exists a node $d \varepsilon N - a$ such that for any other node $c \varepsilon N - a$, exactly one of (cd) and $(dc) \varepsilon N$. Suppose, without loss of generality, $(cd) \varepsilon N$. Then, $N + (ad)$ has degree $(-k)$ and $N + (da)$ has degree $(-k + 1)$, which is a contradiction, so N has no isolated nodes.

Divide the nodes of N into three classes: $X =$ set of initial nodes, $Y =$ set of end nodes, and $Z =$ set of nodes which are both initial and end nodes. Let these sets have q, p, and $m - q - p$ members respectively. It is simple to see that if $q = 0$ or $p = 0$ there is a connected subnetwork of N, which is impossible. Suppose $q \geqq p$.

Since the nodes of X terminate no links, at least q links will have to be added to N to produce a connected supernetwork. We shall now show that q links suffice. There are maximal subsets X_1 and Y_1 such that there is a 1 : 1 correspondence $x_i \varepsilon X_1$, $y_i \varepsilon Y_1$, $i = 1, 2, \cdots, s$, and $(x_i y_i) \varepsilon N$. This follows from the fact that neither X nor Y are empty and from any $x \varepsilon X$ there is either a link to a $y \varepsilon Y$ or a chain via Z to a $y \varepsilon Y$. But in the latter case, $(xy) \varepsilon N$ for if not then N can be connected by the same set of links which connect $N + (xy)$.

The addition of the s links $(y_1 x_2), (y_2 x_3), \cdots, (x_s y_1)$ to N creates a

connected network on the nodes $X_1 + Y_1$. Let $\xi \varepsilon X - X_1$ and $\eta \varepsilon Y - Y_1$; then $(\xi\eta) \notin N$, since X_1, Y_1 are maximal. Thus, if $\xi \varepsilon X - X_1$, there exists $y \varepsilon Y_1$ such that $(\xi y) \varepsilon N$, and if $\eta \varepsilon Y - Y_1$, there exists $x \varepsilon X_1$ such that $(x\eta) \varepsilon N$. Now, from each of the $p - s$ nodes of $Y - Y_1$ introduce links to the nodes of $X - X_1$ such that no two terminate on the same node; this is possible since $q \geqq p$. To each of the remaining nodes of $X - X_1$, if any, introduce a link from a node of Y_1. It is easy to see the resulting network is connected and that $s + (q - s) = q$ links have been added.

If $z \varepsilon Z$ and $y \varepsilon Y$, then $N + (yz)$ still requires the addition of q links to connect, so $Z = 0$ and $p = m - q$. If p were > 1, then for $y_1, y_2 \varepsilon Y$, $N + (y_1y_2)$ would also require the addition of q links to connect; hence $p = 1$ and $q = k + 1$. Thus, $m = q + 1 = k + 2$.

If $p \geqq q$, a similar argument applies.

3. Analogue to Menger's theorem. In graph theory, a node of a connected graph is called a *cut-node* if its removal, along with the incident arcs, results in a graph having two or more components. We generalize this notion: a set of nodes of a connected network is called a *cut-set* if it is one of the smallest sets of nodes whose removal, along with the incident links, results in a disconnected network. If the cut-sets of a network each have κ members, we say the network has *index* κ. It is clear that every connected network has a unique index κ, that $1 \leqq \kappa \leqq m - 1$, and that a connected graph has a cut-node if and only if the index is 1.

The notions of index and degree are parallel with respect to the removal of nodes and links, and so presumedly their values cannot be completely independent. Our first result establishes a relation between them.

THEOREM 2. *Let a connected network on m nodes have degree k and index κ, then $\kappa \leqq k \leqq (m - 1 + \kappa)/2$.*

Proof. To show the left side of the inequality we prove: If a connected network N has degree $k < m - 1$ and (a_jb_j), $j = 1, 2, \cdots, k$, are a set of links whose removal disconnects N, then there exists a set of nodes c_j, $j = 1, 2, \cdots, k$, with $c_j = a_j$ or b_j, such that their removal results in a disconnected network. The c_j are not necessarily distinct.

For $m = 2$ this is obvious.

Consider $m \geqq 3$ and $k = 1$. Let c and d be two nodes such that there is no chain from c to d in $N' = N - (a_1b_1)$. Let M_d consist of d and any node i such that there is a chain from i to d in N', and let $M_c = M - M_d$.

Since $m \geqq 3$ one of these sets has more than one member and neither is empty since $c \varepsilon M_c$ and $d \varepsilon M_d$. If either set, say M_c, has but one member, then the other contains either a_1 or b_1, say b_1. But there is no chain from c to $M_d - b_1$. If both sets have two or more members, remove either a_1 or b_1 and there is no chain between the resulting sets.

For $m \geqq 3$ and $k > 1$ we use an inductive argument. Remove the link $(a_k b_k)$ to obtain N' having degree $k - 1$. By the induction hypothesis there exists a set of no more than $k - 1$ nodes c_j, with $c_j = a_j$ or b_j, $j = 1, 2, \cdots, k - 1$, whose removal from N' results in a disconnected network N''. If either a_k or $b_k \notin N''$ we are done; otherwise, call $N'' + (a_k b_k) = N^*$. Since $k < m - 1$, $m^* \geqq m - (k - 1) \geqq 3$, so if N^* is connected we may apply the $k = 1$ case to show that either the removal of a_k or b_k disconnects N^*. Thus, for $k < m - 1$, $\kappa \leqq k$; however, $\kappa \leqq k$ trivially when $k = m - 1$, so we are done.

We now show the right half of the inequality. If $\kappa = m - 1$, it is trivially true, so we suppose $\kappa < m - 1$. Let S be a cut-set of N and call the resulting disconnected network N'. Let $M' = M - S$. M' consists of more than one node since $m - \kappa > 1$, hence there exist $c, d \varepsilon M'$ such that there is no chain from c to d in N'. Let M_c and M_d be defined as in the first half of the proof. One or the other contains no more than half of the nodes of M', i. e., no more than $(m - \kappa + 1)/2$ nodes. Without loss of generality, we may suppose M_c is the smaller set. In N consider the removal of the links (ci) where $i \varepsilon M_c + S$, which are no more than $(m - \kappa + 1)/2 + \kappa - 1 = (m - 1 + \kappa)/2$ in number. Clearly the resulting network is not connected because there is no link for which c is the initial node, which concludes the proof.

The right inequality is weak and may be improved by relating the degree to the diameter of a network. Let δ_{ab} be the shortest chain from a to b in a connected network, then $\delta = \max_{a,b} \delta_{ab}$ is called the *diameter* of the network.

THEOREM 3. *For a connected network of diameter $\delta > 2$ and degree k, $k \leqq (m - \delta)/2 + 1$.*

Proof. If $\delta = m$, then there is a circuit on the nodes of N such that at least one of the nodes is the initial node of only one link, thus $k = 1 = (m - m)/2 + 1$.

Consider $2 < \delta < m$. Let a and b be two nodes having no chain with fewer than δ links from a to b. If $a \neq b$, there are $\delta + 1$ nodes S in the shortest chain from a to b and $m - \delta - 1$ nodes in $M - S$. If $i \varepsilon M - S$

then not both (ai) and $(ib) \varepsilon N$ since $\delta > 2$. Thus, either a is the initial node of no more than $(m - \delta - 1)/2$ links to $M - S$ or b is the end node of no more than $(m - \delta - 1)/2$ links from $M - S$. Furthermore, a is the initial node of only one link to the nodes of S and b is the end node of only one from S, else there is a chain with fewer than δ links from a to b. Consequently, the removal of at most $(m - \delta - 1)/2 + 1 < (m - \delta)/2 + 1$ links disconnects N.

If $a = b$, S has δ nodes and $M - S$ has $m - \delta$, and by a similar argument $k \leqq (m - \delta)/2 + 1$.

Observe that for $\delta > 2$, Theorem 3 implies the right side of Theorem 2, for $k \leqq (m - \delta + 2)/2 \leqq (m - 1)/2 < (m - 1 + \kappa)/2$.

We turn now to Menger's theorem [3]. It is proved for graphs; however, substantially the same proof holds for networks and so we state it in that form: If a network is connected and has no cut node, i. e., index $\kappa \geqq 2$, then from any node a to any node b there are at least two chains which have no intermediate nodes in common. Because of the parallel definitions of degree and index, one is led to inquire if the following analogue to Menger's theorem is true: If a network has degree $k \geqq 2$, then from any node a to any node b there are at least two chains which have no links in common. It is indeed true; one proof parallels very closely the demonstration given by Dirac for a strengthened form of Menger's theorem; cf. [2], p. 72. We shall not include this proof, for the result is included in the following considerably stronger result.

THEOREM 4. *If a network has degree k, then from any node a to any other node b there is a set of at least k chains such that no two have a common link.*

Proof. We proceed by induction on k; for $k = 1$ the theorem is trivial.

If N has degree $k > 1$, select a k-descendant N' of N (i. e., one of the smallest complete k-minimal subnetworks of N, see p. 705 of (A)). It suffices to show the theorem for N'. Let n be the length of the shortest chain from a to b. If $n = 1$, remove the link (ab) yielding a network of degree $k - 1$, which, by the induction hypothesis, has $k - 1$ chains from a to b with no link in more than one of them. But (ab) is not common to any of them, so there are k chains from a to b in N such that no pair has a common link.

The remainder of the argument is an induction on n with k fixed. Let λ be a chain from a to b of length n and let c be the node of λ immediately preceding b. The shortest chain from a to c has $n - 1$ links, so by the

induction hypothesis there exists a set A_1 of k chains from a to c having no link common to any pair. Similarly, there is a set B of k chains from c to b having no link common to any pair. We may suppose that at least one chain of B has a link in common with a chain of A_1, else we are done.

Notation. If g and h are two nodes of a chain λ, let $\lambda(g, h)$ denote the part of λ from g to h.

Suppose $\beta \in B$ has a link in common with a chain of A_1. Proceed along β opposite to its orientation, i. e., from b toward c, until the first link which is common to a chain, say α, of A_1. Continue further along β until either there is a link common to some $\alpha' \in A_1$, $\alpha' \neq \alpha$, or until c is reached. Let g be the end node of the common link or c, whichever is appropriate. Observe that α and $\beta(g, b)$ may have several common links. Let h be the first node of α, measured along α from a, such that the links of α and $\beta(g, b)$ for which h is the initial node are different. We call $\beta(h, b)$ the *tail* of β.

The remainder of the proof is concerned with the construction of k chains from a to b which satisfy the conditions of the theorem. Parts of chains in A_1 and B will be used. The construction is expedited by dividing A_1 into a number of classes.

A_1 is given. Suppose A_{j-1}, C_{j-1}, D_{j-1}, E_{j-1}, F_{j-1}, and G_{j-1} to be defined. Then define $A_j = D_{j-1} + E_{j-1}$.

Now, for any $\alpha \in A_j$, let $\beta^j{}_\alpha$ be the j-th distinct chain of B as measured along α from a, which has a link in common with α. Let $g^j{}_\alpha$ be the first node in α which is initial to a link of $\beta^j{}_\alpha$ which is not also a link of α. Then we define

$C_j = [\alpha \in A_j \mid \beta^j{}_\alpha(g^j{}_\alpha, b)$ is the tail of $\beta^j{}_\alpha]$.

$D_j = [\alpha \in A_j \mid \alpha \in A_j - C_j, \beta^j{}_\alpha(g^j{}_\alpha, b)$ has a link in common with some $\alpha' \in A_j - C_j, \alpha' \neq \alpha]$.

$E_j = [\alpha \in A_j \mid \alpha \in A_j - C_j, \beta^j{}_\alpha(g^j{}_\alpha, b)$ has links in common only with members of $\sum_{\sigma=1}^{j} C_\sigma + \sum_{\sigma=1}^{j-1} F_\sigma$ and $\beta^j{}_\alpha$ is associated with some $\alpha' \in \sum_{\sigma=1}^{j} C_\sigma + \sum_{\sigma=1}^{j-1} F_\sigma$, by its defining property].

$F_j = [\alpha \in A_j \mid \alpha \in A_j - C_j, \beta^j{}_\alpha(g^j{}_\alpha, b)$ has links in common only with members of $\sum_{\sigma=1}^{j} C_\sigma + \sum_{\sigma=1}^{j-1} F_\sigma$, and $\beta^j{}_\alpha$ is not associated with any $\alpha' \in \sum_{\sigma=1}^{j} C_\sigma + \sum_{\sigma=1}^{j-1} F_\sigma$ by its defining property].

$G_j = [\alpha \in A_j \mid$ no $g^j{}_\alpha$ exists].

Continue this inductive subdivision of A_1 until $A_\eta \neq 0$ and $A_{\eta+1} = 0$.

Let $W_j = [\alpha(a, g^j{}_\alpha)\beta^j{}_\alpha(g^j{}_\alpha, b) \mid \alpha \in C_j + F_j]$. As above, we shall speak of the β's as being associated with the corresponding α's according to the definition of W_j. Now suppose $\omega \in W_i$ and $\omega' \in W_j$ have a link in common. For simplicity we write $\omega = \alpha\beta$, $\omega' = \alpha'\beta'$, and suppose $i \leqq j$. Either β' has a link in common with α or β with α'. Consider the former case. Certainly $\alpha' \notin C_j$ since β' is not a tail, so $\alpha' \in F_j$. But since β' has a link in common with α, then for some $\rho < i$, $\alpha \in D_\rho + E_\rho$, which implies that β' has a link in common with some $\alpha^* \in A_\rho - C_\rho$ or that β' has already been associated with some $\alpha'' \in \sum_{\sigma=1}^{\rho} C_\sigma + \sum_{\sigma=1}^{\rho-1} F_\sigma$. The latter is impossible since $\alpha' \in F_j$. In the former, continue along β' toward b; there is a last chain $\lambda \in A_\rho - C_\rho$ which has a link in common with β'. Either $\lambda \in C_\rho + F_\rho$ or β' has already been associated with a member of A_1, both of which are impossible. So β' and α do not have a common link. For the second case, in which β has a link in common with α', we may suppose $i < j$ since the case $i = j$ has already been covered. Thus, $\alpha \in F_i$, but since $\alpha' \in A_j \subset A_i - C_i$ and β has a link in common with α', then $\alpha \in D_i$ by definition. This is a contradiction.

Let $W' = \sum_{j=1}^{\eta} W_j$ have r members; we have shown there are r chains from a to b such that no two have a link in common.

Let $G = \sum_{j=1}^{\eta} G_j$ have s members, then we show $s + r = k$. If S is a finite set we denote by $N(S)$ the number of elements in S. By definition

$$A_j = C_j + F_j + G_j + A_{j+1},$$

and since the sets on the right are mutually exclusive,

$$\sum_{j=1}^{\eta} N(A_j) = \sum_{j=1}^{\eta} [N(C_j) + N(F_j)] + \sum_{j=1}^{\eta} N(G_j) + \sum_{j=2}^{\eta+1} N(A_j).$$

By choice, $N(A_{\eta+1}) = 0$ and $N(W_j) = N(C_j) + N(F_j)$, so

$$N(A_1) = k = \sum_{j=1}^{\eta} N(W_j) + \sum_{j=1}^{\eta} N(G_j).$$

But, $W_i \cap W_j = 0$, $G_i \cap G_j = 0$, so $N(W') = \sum_{j=1}^{\eta} N(W_j)$ and $N(G) = \sum_{j=1}^{\eta} N(G_j)$, so $k = r + s$.

Let B' be the set of $\beta \in B$ which have no link in common with any $\omega \in W'$. Clearly, $N(B') = N(A_1) - N(W') = k - r = s$. Thus, we may set up an arbitrary 1 : 1 relation between the s elements of G and the s elements of B'.

Denote it by a subscript q and call the set of chains $\alpha_q\beta_q$, $q = 1, 2, \cdots, s$, $\alpha_q \in G$, $\beta_q \in B'$, W''. $W = W' + W''$ is a set of k chains from a to b, which we now show concludes the proof.

First, let $\omega' = \alpha'\beta' \in W'$ and $\omega'' = \alpha''\beta'' \in W''$. By definition of W'', β'' has no link in common with α'. If β' has a link in common with α'', then since $\alpha'' \in G$, β' must have been associated with some $\alpha^* \neq \alpha'$, whence $\alpha'\beta' \notin W'$. Finally, if $\omega = \alpha\beta$ and $\omega' = \alpha'\beta' \in W''$ and, say, β has a link in comon with α', then since $\alpha' \notin \sum_{\sigma=1}^{\eta} C_\sigma + \sum_{\sigma=1}^{\eta} F_\sigma$, β must have been associated with some α^*, whence $\beta \notin B'$. The proof is concluded.

From the analogue to Menger's theorem one may deduce the structure of 2-minimal graphs.

THEOREM 5. *If G is a 2-minimal graph, there exist minimal subnetworks N_1 and N_2 such that $G = N_1 + N_2$ and N_1 and N_2 are (opposite) orientations of G.*

Proof. Let N be a descendant of G and suppose there is an arc $ab \in G - N$. Let α be a chain of N from a to b and β from b to a. There is a link (cd) of α such that (dc) is in β, for otherwise there are two chains of arcs between a and b in $G - ab$ having only nodes in common. Thus $G - (ab)$ has degree $\geqq 2$, which is contrary to assumption. Let $N' = N + (ab) - (cd)$. If N' is connected it is also minimal since it has the same number of links as N. To show it connected it is sufficient to show a chain from b to a and one from c to d. The chain β from b to a remains and $\beta(c, a)(ab)\beta(b, d)$ exists.

If $G - N'$ has an arc, continue the process until N_1 is obtained such that $G - N_1 = N_2$ is arc-free. N_1 is also arc-free, for if $ab \in N_1$ then by Theorem 3. 4 of (A) N_1 consists of two disjoint connected subnetworks joined only by ab. But since G is 2-minimal there is another chain of arcs from a to b not including ab, so N_2 has an arc, a contradiction. Since N_1 is arc-free it is an orientation of G, hence N_2 is a connected orientation of G, and so is minimal.

Finally, it should be observed that Theorem 4, a generalization of a result suggested by Menger's theorem, in turn suggests a generalization to his theorem, to wit: *If a network has index κ, then from any node a to any other node b there exists a set of at least κ chains such that no two have a common intermediate node.* Since the proof of Theorem 4 is based on two sets of chains with a common node c, it is evident that no minor modification of that proof will suffice to demonstrate the above statement, and I have been unable to develop a proof of it.

Some interest attaches in either proving it or giving a counter example, for if it is true there are theorems in graph theory (cf. [2], Theorems 1, 4, 5) of the form "If a graph has no cut-node, then . . ." which presumedly can be strengthened to a form "If a graph has index κ, then . . ."

4. *h*-transitive networks. As was pointed out in the introduction, the conditions of transitivity and connectedness result in the single class of networks, the complete graphs, so it is desirable to weaken the transitivity condition. We shall call a network N *h-transitive* if there is at least one chain $(ab, h) \in N$ such that $a \neq b$, and if for every chain (cd, q) such that $c \neq d$, then $(cd) \in N$ if $q \geqq h+1$ and $(cd) \notin N$ if $1 < q \leqq h$. Clearly, $1 \leqq h \leqq m-2$, and for connected networks, 1-transitivity implies transitivity.

For connected networks, two cases can be distinguished: either there exists a chain of length $\geqq h+1$, or there does not. In the latter case, it is easy to see that the network is minimal. This case has been discussed in (A), so we shall be interested only in the former case.

The following are a set of examples of non-minimal, h-transitive networks with $m \geqq h+2 \geqq 5$. Let Q be a set of four nodes 1, 2, 3, and 4, R a set of $h-2$ nodes distinct from Q labeled $5, 6, \cdots, h+2$, and S a set of $m-h-2$ nodes disjoint from $Q+R$. Let the following links be present on $Q+R+S$: (13), (14), (23), (24), (35), (45), (56), $\cdots$, $(h+1, h+2)$, $(h+2, i)$, $(i3)$, $(i4)$, where $i \in S$. It is not difficult to show these networks satisfy the above requirements.

LEMMA 2. *If N is h-transitive, $h > 1$, and there exists $(ab, q) \in N$ with $q > h$, then $q = h+1$.*

Proof. $(ab, q) = (ac)(cb, q-1)$, and if $q > h+1$, $q-1 > h$, so (cb), $(ab) \in N$. But for $h \geqq 2$, $(ac)(cb) \in N$ implies $(ab) \notin N$, a contradiction.

In (A) a network was called *uniform* if every connected subnetwork has degree 1. A graph which consists of only a circuit of arcs encompassing all the nodes is called a *circle*.

THEOREM 6. *If a network is connected and h-transitive, $h > 1$, then it is uniform or a circle (which is 2-minimal and for which $h = m-2$).*

Proof. If N is minimal, it is uniform (p. 704 of (A)).

If N is non-minimal, there is an $h+1$ chain and, by Lemma 2, it is the longest chain in N. Let its nodes be ordered by the orientation of the chain and $M_1 = \{a, a+1, \cdots a+h, a+h+1 = b\}$ and $M_2 = M - M_1$.

If $h < m - 2$, then $M_2 \neq 0$. If $(ba) \in N$, then a simple induction shows there is a circle on M_1. Then, any link from a node of M_1 to one of M_2 results in an $h+2$-chain, and at least one such link exists since N is connected. By Lemma 2, this is impossible, so $(ba) \notin N$. Let $c \in M_2$, then $(bc) \notin N$ or $(ab, h+1)(bc)$ would be an $h+2$ chain. But since N is connected, there exists at least one $a+i \in M_1$, $1 \leqq i \leqq h$, such that $(b, a+i) \in N$. However, for $j > i$, $(b, a+j) \notin N$ since $(b, a+i)(a+i, a+i+1) \cdots (a+j-1, a+j)$ is a chain of length no greater than $1 + (h-1) = h$. Therefore, b is the initial node of exactly one link, so N has degree 1.

If $h = m - 2$, then $M_2 = 0$. The only possible links to the node $a+h$ are $(b, a+h)$ and $(a+h-1, a+h)$, since any others produce a chain (ab, q) with $q \leqq h$. Thus, the degree of N is, in this case, no greater than 2.

Suppose N is $(m-2)$-transitive and of degree 2, then we show N is a circle (the converse is trivial). Since $k = 2$, $(b, a+h) \in N$. Now node $a+h-1$ must be the end node of at least two links, one being $(a+h-2, a+h-1)$. Of the other two possibilities, $(b, a+h-1)$ and $(a+h, a+h-1)$, the former is excluded because $(b, a+h-1)(a+h-1, a+h)$ implies $(b, a+h) \notin N$, contrary to what we have just shown. Proceed inductively and a circle results.

Now consider the non-minimal h-transitive networks of degree 1. Let S be a connected subnetwork of N, and let h' be the length of the longest chain in S. Either $h' = h+1$ or $h' \leqq h$. In either case, S is h'-transitive, and so the degree of S is 1 except, possibly, if $h' = 1$ or $h' = m' - 2$. If $h' = 1$, then since $h > 1$, $m' = 2$, and so the degree is 1. If $h' = m' - 2$, the only interesting case is degree 2, which, by what we have just seen, implies S is a circle. But, then, $h' = h$, and N is a circle, for $h = m - 2$, else there is an $h+2$ chain. This is contrary to assumption, so S has degree 1, and N is uniform.

The second example on p. 719 of (A) shows there are uniform networks which are not h-transitive.

COROLLARY. *For $m \geqq 5$, there are no 2-transitive, connected, non-minimal networks.*

Proof. Suppose N is 2-transitive, connected, and non-minimal. Let the nodes of one of the 3-chains be a, $a+1$, $a+2$, b. As in the first part of the above proof, if $m \geqq 5$, there is a link from b to $a+i$, $1 \leqq i \leqq 2$. If $(b, a+1) \in N$, then $(ab)(b, a+1)(a+1, a+2) \in N$ implies $(a, a+2) \in N$, which is impossible. Thus, for N to be connected, $(b, a+2) \in N$. If $(a+2, a) \notin N$, then there is a 3-chain from b to a, which is impossible.

But, $(b, a+2)(a+2, a)(a, a+1) \varepsilon N$ implies $(b, a+1) \varepsilon N$, which we have just shown is impossible. Thus, N does not exist.

THEOREM 7. *Let N be connected, non-minimal, h-transitive, $h > 1$, and not a circle. If N contains an arc ab, then N consists of two connected subnetworks N_a and N_b joined only by the arc ab. Either N_a or N_b is h-transitive and non-minimal, and the other is either minimal or h-transitive and non-minimal.*

Proof. Let $M_a = [i \varepsilon M \mid$ there exists (ai, q) not including $b] + a$,

$M_b = [i \varepsilon M \mid$ there exists (bi, q) not including $a] + b$,

$M_{ab} = M_a \cap M_b$,

and

$M'_a = [i \varepsilon M \mid$ there exists (ia, q) not including $b] + a$,

$M'_b = [i \varepsilon M \mid$ there exists (ib, q) not including $a] + b$,

$M'_{ab} = M'_a \cap M'_b$.

Observe that $M = M_a + M_b = M'_a + M'_b$. If $i \varepsilon M_a \cap (M - M'_a)$, then there exist $(ai, r) \varepsilon N$ not containing b and all $(ia, s) \varepsilon N$ do contain b, so there exists a chain α from a to b not including (ab). Since $(ab) \varepsilon N$, which is h-transitive, α must be of length $h+1$. Since $(ba) \varepsilon N$, we may use the same induction as in the proof of Theorem 6 to show there is a circle of arcs on the nodes of α. This contradicts the fact that N is uniform. Thus, $M_a \subset M'_a$. Similarly, $M'_a \subset M_a$, so $M_a = M'_a$. In like manner, $M_b = M'_b$. Thus, $M_{ab} = M_a \cap M_b = M_a \cap M'_b$, so by the same argument $M_{ab} = 0$. Similarly, $M'_{ab} = 0$. Let N_a and N_b be the maximal subnetworks of N on M_a and M_b. By what we have just shown they are connected and they have no node in common. They are joined by ab, and no other link exists between them since $M_{ab} = 0$.

Not both N_a and N_b are minimal, for if they were then N would be minimal. Indeed, no $h+1$ chain traverses the arc ab, for if it did, there would exist another link between N_a and N_b. Thus, one of them is h-transitive and non-minimal, and the other is minimal or h-transitive, non-minimal.

The class of non-minimal, h-transitive, uniform networks on m nodes is smaller than the class of minimal networks on m nodes, and the former can be readily obtained from the latter. Observe, if N is h-transitive and non-minimal, it contains a minimal N' as a descendant. N' is $h+1$-transitive, and N is obtained inductively from N' by introducing a link (ab) every time a chain $(ab, h+1)$ appears. It is easy to find examples of

minimal networks for which this operation does not result in an h-transitive network, so the class of minimal networks is the larger.

For example, if $m = 5$, it is easy to construct the 15 possible minimal networks using Theorem 3.4 of (A). Of these, 10 have arcs and in each case the longest chain in the network passes through the arc, so by Theorem 7 they cannot be descendants of an h-transitive non-minimal network. Of the remaining five, one is the circuit which obviously becomes the circle, and one has $h = 3$ which by the corollary to Theorem 3 cannot yield a 2-transitive case. Performing the inductive operation described above on the other three gives the complete graph in two cases and a 3-transitive network in the third case (which is included in the example at the beginning of this section).

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

REFERENCES.

[1] G. A. Dirac, "Note on the colouring of graphs," *Mathematische Zeitschrift*, vol. 54 (1951), pp. 347-353.

[2] ———, "Some theorems on abstract graphs," *Proceedings of the London Mathematical Society*, ser. 3, vol. 2 (1952), pp. 69-81.

[3] D. Konig, *Theorie der endlichen und unendlichen Graphen*, New York, Chelsea, 1950.

[4] R. D. Luce, "Two decomposition theorems for a class of finite oriented graphs," *American Journal of Mathematics*, vol. 74 (1952), pp. 701-722.

MODULES OVER OPERATOR ALGEBRAS.*[1]

By IRVING KAPLANSKY.

1. Introduction. In [3], Th. 3 the author proved that any *-automorphism of an AW^*-algebra [2] of type I, leaving the center elementwise fixed, is inner. Now in the case of a factor (that is, the algebra of all bounded operators on a Hilbert space) a better result is known, for then *every* automorphism is inner. This leaves a gap that deserves to be filled. A companion problem is the following: is every derivation of an AW^*-algebra of type I inner? It appears that existing AW^* techniques are inadequate to solve these problems, and this paper is devoted to introducing a new technique that does the trick. In brief: the new idea is to generalize Hilbert space by allowing the inner product to take values in a more general ring than the complex numbers. After the appropriate preliminary theory of these AW^*-modules has been developed, one can operate with a general AW^*-algebra of type I in almost the same way as with a factor.

Besides solving the two problems mentioned above, the introduction of AW^*-modules simplifies portions of [3], and also enables us to settle the existence question left open there: we are now able to construct an $\aleph$-homogeneous AW^*-algebra of type I for any prescribed $\aleph$ and center.

In the more special case of W^*-algebras, the two problems could be handled by available tools. But even here AW^*-modules seem to provide the natural method. One may expect that the theory of AW^*-modules will have further applications, both to W^* and to AW^*-algebras.

2. C^*-modules. Let A be a commutative C^*-algebra with unit,[3] and let H be an A-module in the ordinary algebraic sense (including the assumption that the unit element of A acts as unit operator). We shall put the elements of A (typically $a, b, \cdots$) on the left of the elements of H (typically

* Received May 12, 1953.

[1] This paper was prepared with the partial support of the Office of Naval Research.

[2] Definitions will not be repeated from [2] and [3]; but few actual results will be used from the two papers.

[3] The assumption of a unit element is not vital, but it seems pointless to omit it, since A will shortly be an AW^*-algebra. On the other hand, extension of the theory to modules over non-commutative C^*-algebras presents many difficulties.

$x, y, \cdots)$. Suppose there is defined on H an inner product taking values in A and satisfying:

(1) $(x, y) = (y, x)^*$,

(2) $(x, x) \geqq 0$ and is 0 only for $x = 0$,

(3) $(ax + x_1, y) = a(x, y) + (x_1, y)$,

for all x, x_1, y in H and a in A. Of the immediate consequences of the axioms we call explicit attention only to the following: $(x, ay) = a^*(x, y)$, and we pass on at once to the introduction of a norm. As compared with Hilbert space, there are in fact two "norms" available, one A-valued and the other numerical; they both have a role to play in our work. We use the notation

$$| x | = (x, x)^{\frac{1}{2}}, \qquad \| x \| = \| (x, x) \|^{\frac{1}{2}},$$

where on the right we mean the usual positive square root and norm in A. We have that $\| x \|$ is the norm of $| x |$ in the algebra A; alternatively, $\| x \|$ is the sup of $| x |$ when the latter is regarded as a function on the space of maximal ideals in A.

The Schwarz inequality

$$| (x, y) | \leqq | x | \, | y | \tag{1}$$

can be verified by adapting a standard proof. It is also possible to reduce to the numerical case by the following device. For any maximal ideal M in A we may define a numerical inner product $(x, y)(M)$ on H, by mapping modulo M. In this inner product there may exist non-zero elements x with $(x, x)(M) = 0$; nevertheless the Schwarz inequality

$$| (x, y))(M) | \leqq | x | (M) | y | (M) \tag{2}$$

is known to hold. The validity of (2) for every M is precisely equivalent to (1). On taking norms in (1) we further get the numerical version of the Schwarz inequality:

$$\| (x, y) \| \leqq \| x \| \, \| y \|. \tag{3}$$

From (1) or (3) we deduce in the usual way the triangle inequality $\| x + y \| \leqq \| x \| + \| y \|$, and we have that H is a normed linear space. By (3) the inner product is jointly continuous in its arguments; likewise the equality $| ax | = | a | \, | x |$ shows that ax is jointly continuous in its arguments. So H may be completed with preservation of all the postulates. If H is already complete, we shall call it a *C*-module over A*. Thus a *C**-

module is a kind of blend of a Hilbert space and a commutative C^*-algebra.

By a *bounded operator* T from a C^*-module H into a second one K we mean a mapping of H into K which is not only linear and continuous as usual, but also a *module homomorphism*. Thus (if we place T on the right of the elements of H) we have $(ax)T = a(xT)$ for all a and x. The set B of all bounded operators on H forms a Banach algebra in the usual operator norm.

We call T^* the *adjoint* of T if $(xT, y) = (x, yT^*)$ for all x and y. The question of the existence of T^* will be discussed in § 8 only with the aid of more axioms. But whenever T^* exists, we can verify the equation $\| TT^* \| = \| T \|^2$.

LEMMA 1. *Let H be a C^*-module, and T a bounded operator on H with an adjoint T^* which is also a bounded operator. Then $\| T \| = \| T^* \|$ and $\| TT^* \| = \| T \|^2$.*

Proof. We have

$$| xT |^2 = (xT, xT) = (xTT^*, x) \leqq | x | \, | xTT^* |, \tag{4}$$

the last step by the Schwarz inequality (1). We take norms in (4), recalling that $\| T \|$ is the sup of $\| xT \|$ for $\| x \| \leqq 1$. The result is the first inequality in

$$\| T \|^2 \leqq \| TT^* \| \leqq \| T \| \, \| T^* \|, \tag{5}$$

while the second inequality holds in any Banach algebra. On cancelling $\| T \|$ we get $\| T \| \leqq \| T^* \|$. Since T is also the adjoint of T^*, the reverse inequality also holds. Hence $\| T \| = \| T^* \|$, and insertion of this into (5) completes the proof.

3. *AW-modules.** It does not appear to be possible to get much deeper into the subject without imposing further postulates, acting as a sort of algebraic substitute for a weak topology. These postulates are motivated by two properties of an AW^*-algebra A ([2], Lemmas 2. 2 and 2. 5).

(a) Let $\{e_i\}$ be orthogonal projections in A with l. u. b. e, and suppose a is an element of A with $e_i a = 0$ for all i; then $ea = 0$.

(b) Let $\{e_i\}$ be central orthogonal projections in A with l. u. b. 1, and let $\{a_i\}$ be a bounded subset of A; then there exists in A an element a with $e_i a = e_i a_i$ for all i.

We propose to assume outright the analogues of these two properties.

Definition. Let A be a commutative AW^*-algebra. We say that H is an *AW^*-module over A* if it is a C^*-module over A and further enjoys the following two properties:

(a) Let $\{e_i\}$ be orthogonal projections in A with l. u. b. e, and suppose x is an element of H with $e_i x = 0$ for all i; then $ex = 0$.

(b) Let $\{e_i\}$ be orthogonal projections in A with l. u. b. 1, and let $\{x_i\}$ be a bounded subset of H; then there exists in H an element x with $e_i x = e_i x_i$ for all i.

It follows from postulate (a) that the element x of (b) is unique, and we shall write $x = \sum e_i x_i$. This is in accordance with the notation $\sum e_i a_i$ used in [3] for the analogous element constructed in the algebra A. These two infinite "sums" are related in the desired way.

LEMMA 2. *Let A be a commutative AW^*-algebra, and $\{e_i\}$ a set of orthogonal projections in A with* l. u. b. 1. *Let H be an AW^*-module over A, y an element of H, and $\{x_i\}$ a bounded subset of H. Then*

$$(\sum e_i x_i, y) = \sum e_i (x_i, y). \tag{6}$$

Proof. It is to be observed that the elements (x_i, y) are bounded in A, and so the right side of (6) is well defined. To prove (6) it is enough to verify that it holds after multiplication by a fixed e_j (we shall be making repeated use of this observation throughout the paper). Then the right side becomes $e_j(x_j, y)$, while the left side is

$$e_j(\sum e_i x_i, y) = (e_j \sum e_i x_i, y) = (e_j x_j, y) = e_j(x_j, y).$$

LEMMA 3. *Let H be an AW^*-module over A. Then the annihilator in A of any subset of H is a direct summand of A.*

Proof. Let I be the annihilator in question; I is an ideal by algebra, and is closed in the norm of A by the continuity of the module operations. Let $\{e_i\}$ be a maximal set of orthogonal projections in I. By postulate (a), their l. u. b. e is again in I. We claim that $I = eA$, and it suffices to prove $I = eI$, or equivalently $(1 - e)I = 0$. If on the contrary $(1 - e)I$ is non-zero, it contains a non-zero projection, which could be used to enlarge the set $\{e_i\}$.

In particular, the annihilator of all of H is a direct summand of A, and we say that H is *faithful* if its annihilator is 0. As a rule, there is no loss of generality in restricting our attention to faithful AW^*-modules.

LEMMA 4. *Let x be a non-zero element of an AW^*-module H over A. Then there exists in A a non-zero projection e and an element a with $a \mid x \mid = e$.*

Proof. We have that $\mid x \mid$ is a non-zero element of the commutative AW^*-algebra A. Consequently there exists a non-zero projection e such that the function representing $\mid x \mid$ is bounded away from zero in the direct summand eA. We take a to be the inverse of $\mid x \mid$ in eA.

LEMMA 5. *A faithful AW^*-module H contains an element x with $\mid x \mid = 1$.*

Proof. Lemma 4 provides us with an element x_1 with $\mid x_1 \mid$ equal to a non-zero projection e_1. We apply Zorn to get a maximal collection $\{x_i\}$ with $\mid x_i \mid = e_i$, and $\{e_i\}$ orthogonal projections, say with l. u. b. e. If $e \neq 1$, we have $(1 - e)H \neq 0$ since H is faithful. Applying Lemma 4 to a non-zero element of $(1 - e)H$, we enlarge the collection $\{x_i\}$. Hence e must be 1. Since each $\mid x_i \mid = 1$, we may form the element $x = \sum e_i x_i$. For any e_j we have

$$e_j(x, x) = (e_j x, e_j x) = (e_j x_j, e_j x_j) = e_j.$$

Thus $(x, x) - 1$ annihilates every e_j and is 0. Hence x is the desired element.

It is appropriate to observe that Lemma 5 may fail if H is merely a C^*-module (even over an AW^*-algebra). For instance take A to be the algebra of all bounded sequences of complex numbers, and H the C^*-module of all sequences aproaching 0.

We now introduce some terminology, imitating the example of Hilbert space. We say that x and y are *orthogonal if* $(x, y) = 0$. The *orthogonal complement* R' of a subset R of H is the set of all x with $(R, x) = 0$. A set $\{x_\lambda\}$ is *orthonormal* if each $\mid x_\lambda \mid = 1$ and $(x_\lambda, x_\mu) = 0$ for $\lambda \neq \mu$. The existence of a maximal orthonormal set is assured by Zorn's lemma. But this is not what we generally need. Rather we want what we shall call an orthonormal basis.

Definition. An *orthonormal basis* in an AW^*-module is an orthonormal set whose orthogonal complement is 0. An AW^*-module is said to be *homogeneous* if it possesses an orthonormal basis, or more precisely it is $\aleph$-homogeneous if it possesses an orthonormal basis of $\aleph$ elements.[4]

---

[4] This terminology is not meant to suggest that the $\aleph$ in question is unique. As regards this problem, we have not advanced beyond what was shown in [3]: $\aleph$ is unique

Not every AW^*-module has an orthonormal basis, but we can split it into homogeneous parts which do. To accomplish this we next launch the theory of submodules.

4. Submodules. The appropriate concept for our purposes is embodied in the following definition.

Definition. Let H be an AW^*-module over A. By an *AW^*-submodule* S we mean a subset satisfying: (1) S is a submodule of H in the ordinary algebraic sense, (2) S is closed in the norm topology of H, (3) if $\{x_i\}$ is a bounded subset of S, and $\{e_i\}$ are orthogonal projections in A with l. u. b. 1, then the element $\sum e_i x_i$ is again in S.

We remark: (a) an AW^*-submodule is itself an AW^*-module over A, (b) the intersection of any number of AW^*-submodules is again an AW^*-submodule, (c) consequently for any subset R there exists a smallest AW^*-submodule containing R; we call it the AW^*-submodule generated by R.

The next two lemmas provide us with two natural sources of AW^*-submodules.

LEMMA 6. *Let H be an AW^*-module. Then the orthogonal complement of any subset of H is an AW^*-submodule of H.*

Proof. The only point worthy of note is this: given that $(x_i, y) = 0$ for every i, prove that $(\sum e_i x_i, y) = 0$. This follows at once from Lemma 2.

LEMMA 7. *Let T be a bounded operator from an AW^*-module H into a second one. Then the kernel N of T is an AW^*-submodule of H.*

Proof. We content ourselves with proving that $x_i \in N$ implies $\sum e_i x_i \in N$. It is enough to prove $(\sum e_i x_i)T = 0$ after multiplication by a fixed e_j, after which it becomes $(e_j x_j)T$, which does vanish.

Now consider a faithful AW^*-module H, a maximal orthonormal set $\{x_\lambda\}$ in H, and its orthogonal complement S. By Lemma 6, S is an AW^*-submodule of H. If S is faithful we can use Lemma 5 to enlarge the set $\{x_\lambda\}$. Hence there must exist a non-zero projection e in A with $eS = 0$. We now claim that the elements $\{ex_\lambda\}$ constitute an orthonormal basis for eH, regarded

---

if A satisfies the countable chain condition locally. The author conjectures that the uniqueness may fail otherwise.

as an AW^*-module over eA. For suppose the element y in eH is orthogonal to all ex_λ. Then

$$0 = (ex_\lambda, y) = e(x_\lambda, y) = (x_\lambda, ey).$$

Hence $ey = y$ is in S, and this implies $y = 0$, since $eS = 0$. When pursued by transfinite induction, the process yields the following theorem.

THEOREM 1. *Let H be a faithful AW^*-module over A. Then there exist in A orthogonal projections $\{e_i\}$ with* l. u. b. 1 *such that each e_iH is a homogeneous AW^*-module over e_iA.*

5. Construction of AW^*-modules. It is time for us to consider the question of the existence of AW^*-modules. Now our fundamental example is A itself, regarded as a module over A. There is no difficulty in discussing the direct sum of a finite number of copies of A. But the construction of an "infinite direct sum" requires more elaborate discussion. While we are at it, we might as well construct the direct sum of arbitrary AW^*-modules over A.

Our fundamental tool for this purpose is the use of the infinite sums in A which were briefly discussed in § 4 of [3]. We recall that the self-adjoint elements of a commutative AW^*-algebra A form a conditionally complete lattice, that is, every bounded set has a least upper bound. If elements $a_\lambda \geqq 0$ are given in A such that there is a fixed upper bound to all finite sums, we say that $\sum a_\lambda$ converges and we define the sum to be the least upper bound of these finite sums. There should be no danger of confusion between these sums and the other kind of infinite sum $\sum e_ix_i$ which we are also using; in particular we shall always use Greek subscripts for the former and Latin for the latter.

The reader should be warned of a possible pitfall: if we think of a_λ as a function on the space of maximal ideals of A, then $\sum a_\lambda$ is not the pointwise sum of these functions (although the two sums differ only on a set of the first category).

We shall further need to make use of sums of elements which are not necessarily self-adjoint or positive. Only absolute convergence is relevant, and so we define $\sum a_\lambda$ to be convergent if $\sum |a_\lambda|$ is. The actual value of the sum $\sum a_\lambda$ is assigned by splitting a_λ into four parts (first into real and imaginary parts, then each of these into positive and negative parts). We shall not pause over routine facts needed in manipulating these sums, samples of which are

$$|\sum a_\lambda| \leqq \sum |a_\lambda|, \qquad b \sum a_\lambda = \sum ba_\lambda.$$

However a crucial rôle is played by a suitable Schwarz inequality.

LEMMA 8. *If* $\sum |a_\lambda|^2$ *and* $\sum |b_\lambda|^2$ *converge, so does* $\sum a_\lambda b_\lambda$, *and*

$$|\sum a_\lambda b_\lambda|^2 \leqq \sum |a_\lambda|^2 \sum |b_\lambda|^2. \tag{7}$$

Proof. We first prove (7) for finite sums, by adapting a standard proof, or by arguing modulo maximal ideals of A. Then, holding a finite sum fixed on the left of (7), we may put in the infinite sums on the right. Since the resulting inequality holds for every finite sum on the left of (7), we get the convergence of $\sum a_\lambda b_\lambda$ and the desired inequality.

Now let an index set I be given, and for each $\lambda \in I$ an AW^*-module H_λ over A. Define H to be the set of all arrays $x = \{x_\lambda\}$ with $x_\lambda \in H_\lambda$, and $\sum |x_\lambda|^2$ convergent. For $x = \{x_\lambda\}$ and $y = \{y_\lambda\}$ in H define (x, y) to be $\sum (x_\lambda, y_\lambda)$. Since $|(x_\lambda, y_\lambda)| \leqq |x_\lambda|\,|y_\lambda|$, the convergence of $\sum (x_\lambda, y_\lambda)$ is assured by Lemma 8. We pass rapidly over the fact that H is in a natural way a C^*-module over A, the verification being routine except for the completeness of H in its norm. This we prove in (c) below, after we have checked the two AW^* postulates.

(a) Let there be given $x = \{x_\lambda\}$ in H, and orthogonal projections e_i in A with l. u. b. e. Suppose each $e_i x = 0$. We have to prove $ex = 0$. Now $e_i x_\lambda = 0$, whence $ex_\lambda = 0$, $ex = 0$.

(b) Let there be given elements $x(i) = \{x_\lambda(i)\}$ in H, with $\|x(i)\|^2$ bounded by a constant K. (Here i of course runs over a second independent index set). Let e_i be orthogonal projections in A with l. u. b. 1. We must construct an element y in H with $e_i y = e_i x_i$ for all i. Now for fixed λ we have $|x_\lambda(i)|^2 \leqq K$. Hence in the AW^*-module H_λ we may form the element $y_\lambda = \sum e_i x_\lambda(i)$, satisfying $e_i y_\lambda = e_i x_\lambda(i)$ for all i. When λ is restricted to a finite subset J of the index set I, we have

$$\sum_{\lambda \in J} |y_\lambda|^2 = \sum_{\lambda \in J} \sum_i e_i |x_\lambda(i)|^2. \tag{8}$$

When (8) is multiplied by e_j, the right side becomes

$$\sum_{\lambda \in J} e_j |x_\lambda(j)|^2$$

and is bounded by $\|x(j)\|^2 \leqq K$. Since this is true for every j, the left side of (8) is likewise bounded by K. Hence $y = \{y_\lambda\}$ is in H. Moreover

$$e_i y = \{e_i y_\lambda\} = \{e_i x_\lambda(i)\} = e_i x_i,$$

and y is the desired element.

(c) Let $x(m) = \{x_\lambda(m)\}$, with m running over the positive integers, be elements constituting a Cauchy sequence in H (relative to the norm in H). We shall produce in H a limit y for this sequence. We begin by noting that for each fixed λ, the elements $x_\lambda(m)$ form a Cauchy sequence in H_λ, converging say to y_λ. The numbers $\| x(m) \|^2$ are bounded, say by K. We claim that also $\sum | y_\lambda |^2 \leqq K$. It is enough to verify this for a finite sum, taken over a finite subset J of the index set I. Now

$$\sum_{\lambda \in J} | y_\lambda |^2 = \lim_{m \to \infty} \sum_{\lambda \in J} | x_\lambda(m) |^2,$$

while for each m

$$\sum_{\lambda \in J} | x_\lambda(m) |^2 \leqq \| x(m) \|^2 \leqq K.$$

Hence $\sum | y_\lambda |^2 \leqq K$, which means that $y = \{y_\lambda\}$ is in H. It remains to prove that $x(m)$ converges to y in the norm of H. Given $\epsilon > 0$, cut in far enough so that all $\| x(m) - x(n) \|^2 \leqq \epsilon$. We claim that

$$\sum_{\lambda} | x_\lambda(m) - y |^2 \leqq \epsilon.$$

For (again) it is enough to verify this with λ restricted to a finite subset J. But

$$\sum_{\lambda \in J} | x_\lambda(m) - y_\lambda |^2 = \lim_{n \to \infty} \sum_{\lambda \in J} | x_\lambda(m) - x_\lambda(n) |^2 \leqq \epsilon.$$

Hence $\| x(m) - y \|^2 \leqq \epsilon$, and this proves the convergence of $x(m)$ to y.

We have thus completed the proof that H is an AW^*-module. Let us now specialize to the case where each H_λ is isomorphic to A. Then H has an evident orthonormal basis: the elements which are 1 at a designated coordinate λ and 0 everywhere else. The cardinal number of this orthonormal basis is the same as the cardinal number of the index set I, and this is at our disposal. We have proved:

THEOREM 2. *For any commutative AW^*-algebra A and cardinal number $\aleph$, there exists an $\aleph$-homogeneous AW^*-module over A.*

6. Orthogonal decomposition. We wish now to prove that any AW^*-submodule is a direct summand, and also that a homogeneous AW^*-module is determined in a suitable way by an orthonormal basis. The following lemma provides the basic information needed in proving both of these facts.

LEMMA 9. *Let H be an AW^*-module over A, let $\{x_\lambda\}$ be an orthonormal set in H, let S be the AW^*-submodule of H generated by $\{x_\lambda\}$, and let $\{c_\lambda\}$*

be elements in A such that $\sum |c_\lambda|^2$ is convergent. Then there exists in S an element t satisfying $|t|^2 = \sum |c_\lambda|^2$, $(t, x_\lambda) = c_\lambda$ for all λ.

Proof. The proof is very much the same as that of Lemma 7 in [3], and consequently we shall not give full details.

Write $w_\lambda = |c_\lambda|^2$, $w = \sum w_\lambda$. Let an integer m be given. We apply [3], Lemma 5, to obtain a set $\{e_i\}$ of orthogonal projections in A with l. u. b. 1, and for each i a certain finite sum v_i of w_λ's such that

$$\| e_i(w - v_i) \| < m^{-2}. \tag{9}$$

Write u_i for $\sum c_\lambda x_\lambda$, taken over the same finite set of λ's as were used in forming v_i. We have $|u_i| \leqq w$, and so we may define t_m in H by $t_m = \sum e_i u_i$. The proof that $\{t_m\}$ is a Cauchy sequence (in the norm topology for H) does not differ materially from the corresponding portion of the proof of [3], Lemma 7, and we omit it. Let t denote the limit of t_m as $m \to \infty$. We note that $u_i \in S$, $t_m \in S$, and so $t \in S$.

It remains to prove that t has the two properties claimed for it in the lemma. We consider (u_i, x_λ) and note that it is c_λ, if λ is one of the subscripts occurring in the defining sum $u_i = \sum c_\lambda x_\lambda$, and is 0 otherwise. In the latter case $e_i |c_\lambda|^2$ is bounded by $e_i(w - v_i)$, and so $|e_i c_\lambda| < 1/m$ by (9). Also

$$(t_m, x_\lambda) = (\sum e_i u_i, x_\lambda) = \sum e_i (u_i, x_\lambda)$$

by Lemma 2. From this it follows that

$$|(t_m, x_\lambda) - c_\lambda| < 1/m$$

always holds. Proceeding to the limit as $m \to \infty$, we deduce $(t, x_\lambda) = c_\lambda$. Again

$$e_j |t_m|^2 = e_j(\sum e_i u_i, \sum e_i u_i) = e_j |u_j|^2 = e_j v_j.$$

Since this holds for all j we deduce from (8) that $w - |t_m|^2 < m^{-2}$. Hence $|t|^2 = w$, and this completes the proof of Lemma 9.

LEMMA 10. *Let H be an AW^*-module, and T a homogeneous AW^*-submodule, with orthonormal basis $\{x_\lambda\}$. Then $\{x_\lambda\}$ and T have the same orthogonal complement in H. Also, for any x in T, we have $|x|^2 = \sum |(x_\lambda, x)|^2$.*

Proof. Given $(x_\lambda, y) = 0$ for all λ, we have to prove $(T, y) = 0$. We take $x \in T$, write $(x, x_\lambda) = c_\lambda$, and observe that $\sum |c_\lambda|^2$ converges (and is in fact bounded by $\|x\|^2$). Let t be the element given us by Lemma 9. Then

t lies in the AW^*-submodule generated by $\{x_\lambda\}$, and a fortiori lies in T. We have that $x - t$ is orthogonal to every x_λ. From the definition of an orthonormal basis (as an orthonormal set whose orthogonal complement is 0), it follows that $x - t = 0$. Again, the orthogonal complement of y is an AW^*-submodule by Lemma 6; it contains $\{x_\lambda\}$, hence contains the AW^*-submodules generated by $\{x_\lambda\}$, hence contains $t = x$. We have proved $(x, y) = 0$ and $(T, y) = 0$. Moreover by Lemma 9 we also have

$$| x |^2 = \sum | c_\lambda |^2 = \sum | (x_\lambda, x) |^2.$$

THEOREM 3. *Let H be an AW^*-module, T an AW^*-submodule, T' its orthogonal complement. Then $H = T \oplus T'$.*

Proof. The problem is to express an arbitrary element z of H as a sum of elements in T and T'. We shall first make the additional assumption that T is homogeneous, say with orthonormal basis $\{x_\lambda\}$. Write $(z, x_\lambda) = c_\lambda$ and observe that $\sum | c_\lambda |^2$ is convergent, being in fact bounded by $\| z \|^2$. Let t be the element of T provided by Lemma 9, with $(t, x_\lambda) = (z, x_\lambda)$ for all λ. By Lemma 10, $z - t$ is in T'. Thus $z = t + (z - t)$ is the desired decomposition.

Next we observe that we may assume that T is faithful. For if eA is the annihilator of T, we perform the decomposition $H = eH + (1 - e)H$, place eH in T' and work inside $(1 - e)H$. Finally we apply Theorem 1 to obtain a set $\{e_i\}$ of orthogonal projections in A with l. u. b. 1, and such that each e_iT is a homogeneous AW^*-module over e_iA. We may then decompose e_iz into a sum $x_i + y_i$ of elements in e_iT and e_iT'. Since $\| x_i \| \leqq \| z \|$, the elements x_i may be strung together to form x in T; similarly the elements y_i yield y in T', and $z = x + y$.

THEOREM 4. *Let A be a commutative AW^*-algebra. Then any two $\aleph$-homogeneous AW^*-modules over A are isomorphic.*

Proof. Let H and K be the modules, $\{x_\lambda\}$ and $\{y_\lambda\}$ their respective orthonormal bases. For x in H we set $c_\lambda = (x, x_\lambda)$ and then by Lemma 9 find y in K with $(y, y_\lambda) = c_\lambda$. By Lemmas 9 and 10 the mapping $x \to y$ is one-to-one, onto, and preserves the norm $| x |$. By the usual polarization argument, the mapping also preserves the inner product. Since it is evidently a module isomorphism, the theorem is proved.

7. Functionals. Let H be an AW^*-module over A. As is appropriate in the present context, we define a *functional* on H to be a module homo-

morphism of H into A. We shall devote the present section to showing that H is self-dual in the same way that Hilbert space is.

We begin with the easy observation that elements of H give rise to functionals in the appropriate way.

LEMMA 11. *Let x be an element of an AW^*-module H. Then the functional on H given by $y \to (y, x)$ has precisely norm $\| x \|$.*

Proof. That the norm is at most $\| x \|$ follows from the Schwartz inequality. On the other hand, the particular case $y = x$ shows that the norm is at least $\| x \|$.

We next note a boundedness criterion that will be used later. With the aid of Lemma 11, it is an immediate consequence of Banach's uniform boundedness theorem ([1], p. 80).

LEMMA 12. *Let R be a subset of an AW^*-module H, and suppose that for every y in H the set (R, y) is bounded. Then R is bounded.*

THEOREM 5. *Let H be an AW^*-module, f a continuous function on H. Then there exists a (unique) element x in H such that*

$$f(y) = (y, x) \tag{10}$$

for all y in H.

Proof. Let N be the kernel of f; by Lemma 7, N is an AW^*-submodule of H. We apply the decomposition of Theorem 3: $H = N \oplus N'$. It is enough to prove Theorem 5 for the functional f restricted to N'. For if we find $x \in N'$ satisfying (10) for all y in N', (10) will automatically be fulfilled also for y in N. Thus, after a change of notation we can assume that f is faithful on H.

We shall next reduce the problem to the case where H is faithful. For let eA be the annihilator of H. We have $ef(y) = f(ey) = 0$ for all y in H, that is, the range of f is automatically in $(1 - e)A$. So we may as well consider f as a functional on H with the latter regarded as a faithful module over $(1 - e)A$.

It being now assumed that H is faithful, we apply Lemma 5 to get an element z with $(z, z) = 1$. The next step of the argument is to prove that the orthogonal complement of z is 0. Suppose on the contrary that $(z, w) = 0$ with $w \neq 0$. We shall derive a contradiction by two suitable applications of Lemma 4. The first application is to the element w, and yields a non-

zero projection e_1 such that $e_1 \mid w \mid$ is a regular element of e_1A. We have $\mid e_1z \mid = e_1 \neq 0$. Since f is faithful, $f(e_1z) \neq 0$. We drop down to a projection $e_2 \leqq e_1$ such that $f(e_2z)$ is a regular element of e_2A. Observe that e_2w cannot be 0. There exists an element b such that $bf(e_2z) = e_2f(w)$, whence $f(be_2z - e_2w) = 0$, $be_2z = e_2w$. On taking inner products with e_2w in this last equation we get 0, since z and w are orthogonal. This yields the contradiction $e_2w = 0$.

To complete the proof of Theorem 5 we set $x = f(z)^*z$. For any y in H we have that $y - (y, z)z$ is orthogonal to z, and, by the preceding paragraph, vanishes. Hence $f(y) = (y, z)f(z) = (y, x)$, as desired.

8. Existence of the adjoint. We are now in a position to establish that the existence of the adjoint of an operator is equivalent to continuity.

THEOREM 6. *Let H be an AW^*-module and T a module homomorphism of H into itself. Then T is continuous if and only if it has an adjoint.*

Proof. Suppose that T is continuous. For fixed z in H the mapping $y \to (yT, z)$ is a continuous functional on H. By Theorem 5 this functional is induced by an element of H, and this element is our choice for zT^*. That T^* is the desired adjoint of T is subject to direct verification.

Suppose that T has an adjoint T^*. To prove that T is continuous it suffices to show that T is bounded on the unit sphere R of H. By Lemma 12 it is enough to know that RT is bounded when the inner product is taken with a fixed element y. But the equation $(RT, y) = (R, yT^*)$ gives us the explicit bound $\| yT^* \|$ for $\| (RT, y) \|$.

The algebra of bounded operators on H is thus a Banach algebra possessing a *-operation with the usual algebraic properties, and Lemma 1 tells us that the equation $\| TT^* \| = \| T \|^2$ is satisfied. It has recently been proved that these properties suffice to make B a C^*-algebra (to be sure, it would not be difficult here to verify directly that $1 + TT^*$ has an inverse).

We shall shortly prove that B is actually a very special kind of C^*-algebra, namely an AW^*-algebra of type I; and every AW^*-algebra of type I arises in this way.

9. Abelian projections. Let B be the algebra of bounded operators on an AW^*-module H. We call an element E in B a projection if it is a self-adjoint idempotent: $E^2 = E$, $E^* = E$. A projection gives rise to the decomposition $HE \oplus H(1 - E)$ of H into the range and null space of E, and

conversely to every AW^*-submodule of H there corresponds in this way a projection (Theorem 3). We say that E is *abelian* if EBE is commutative, and in the next lemma we give a precise determination of the abelian projections.

LEMMA 13. *Let H be an AW^*-module over A. Let y be an element of H with $|y|$ a projection in A. Then the operator E defined by $xE = (x, y)y$ is an abelian projection. Conversely every abelian projection arises in this way.*

Proof. Since $|y|$ is a projection, we have $(y, y)y = y$. It follows that $xE^2 = xE$, so that E is an idempotent. We compute

$$(xE, z) = (x, zE) = (x, y)(y, z),$$

$$xETEUE = (x, y)(yT, y)(yU, y)y = xEUETE$$

for any operators T, U. Hence E is an abelian projection.

Conversely suppose that E is an abelian projection. HE is an AW^*-submodule of H; if we write its annihilator as $(1 - e)A$, then HE is faithful over eA, and by Lemma 5 we may find an element y in HE with $|y| = e$. It should be noted that e acts as unit element on HE. We propose to show that $xE = (x, y)y$ for all x in H. Write $z = xE - (x, y)y$, and observe that $ez = zE = z$. Also

$$(z, y) = (xE, y) - (x, y)(y, y) = (x, yE) - e(x, y) = 0.$$

We define the operators T and U on H by

$$wT = (w, y)z, \qquad wU = (w, z)y.$$

We verify readily that $ETE = T$, $EUE = U$. Since E is abelian, T and U commute. Hence

$$wTU = uUT = (w, y)(z, z)y = (w, z)(y, y)z = (w, z)ez = (w, z)z.$$

On setting $w = z$ and recalling that $(z, y) = 0$, we find that $z = 0$. This completes the proof of the lemma.

10. The algebra of operators. The next two theorems constitute the main results of the paper.

THEOREM 7. *Let A be a commutative AW^*-algebra, H a faithful AW^*-module over A. Let B be the algebra of bounded operators on H. Then B*

is an AW-algebra of type I with center isomorphic to A. If H is $\aleph$-homogeneous, so is B.*

Proof. We begin by identifying the center of B. Multiplication of H by an element of A obviously gives rise to a center element; we have to show that this exhausts the center. Let then T be a central operator. We impose the condition that it commutes with the operator $x \to (x, y)z$. The result is $(xT, y)z = (x, y)zT$. We take x with $(x, x) = 1$ by Lemma 5, and $y = x$. Then $zT = (xT, x)z$, showing that T is simply multiplication by (xT, x).

Next we shall show that B is an AW^*-algebra (it being recalled that we already know B to be a C^*-algebra). For this purpose the official axioms in [2] turn out to be slightly inconvenient. A better axiom is the following: the right annihilator in B of any subset of B is of the form EB with E a projection. (It is implicit in the discussion of § 7 of [2] that this axiom characterizes AW^*-algebras among C^*-algebras). Now the right annihilator of any element of B coincides with the annihilator of its range. So the problem comes to this: given a subset R of H, prove that the annihilator of R is of the form EB. Since the kernel of any bounded operator is an AW^*-submodule (Lemma 7), the annihilator of R coincides with the annihilator of the AW^*-submodule (say R_1) generated by R. Form the orthogonal decomposition $H = R_1 \oplus R_2$ (Theorem 3), and define E to be 0 on R_1, the identity on R_2. Then E is a projection, and the annihilator of R_1 is manifestly EB.

Lemma 13 shows that B possesses abelian projections in abundance, and the same is plainly true for any direct summand of B. Hence B is an AW^*-algebra of type I.

Finally suppose that H is $\aleph$-homogeneous, say with orthonormal basis $\{x_\lambda\}$. Define operators E_λ, $E_{\lambda\mu}$ by

$$zE_\lambda = (z, x_\lambda)x_\lambda, \qquad zE_{\lambda\mu} = (z, x_\lambda)x_\mu.$$

By Lemma 13, the E_λ's are abelian projections, and they are plainly orthogonal. Since no non-zero element of H is annihilated by all the x_λ's, it follows that no non-zero element is annihilated by all the E_λ's, whence the l. u. b. of the E_λ's is the identity operator. Finally $E_{\lambda\mu}E_{\mu\lambda} = E_\lambda$, and so the E_λ's are equivalent. This shows that B is $\aleph$-homogeneous and completes the proof of Theorem 7.

THEOREM 8. *Let B be an AW*-algebra of type I, and e an abelian*

projection in B, not annihilated by any non-zero central element of B.[5] *Then eB is in a natural way a faithful AW^*-module over eBe, and when B is represented by right multiplication on eB, it gives rise to precisely all bounded (module) operators on eB.*

Proof. Since e is abelian, eBe is a commutative AW^*-algebra, and so is eligible to admit AW^*-modules. We define the inner product on eB as $(ex, ey) = exy^*e$. It is then routine to verify that eB is indeed an AW^*-module over eBe. The action of eBe on eB coincides with that of the center of B ([3], Lemma 10), and hence eB is a faithful AW^*-module. The right annihilator of eB within B is generated by a central projection ([2], Th. 2. 3, Cor. 1) and so is 0. Thus the elements of B are faithfully represented by right multiplication on eB. That these right multiplications are bounded (module) operators on eB is clear. It only remains to be seen that B coincides with the full algebra (say B_1) of bounded operators on eB; we recall that by Theorem 7, B_1 is also an AW^*-algebra of type I. The proof that $B = B_1$ will be carried out in two steps, the first of which is to show that B contains all abelian projections in B_1. Now by Lemma 13, any abelian projection in B_1 has the form $ex \to (ex, ey)ey = exy^*ey$ for a certain element ey in eB. But right multiplication by y^*ey has exactly the same effect. Hence this abelian projection is already in B.

The final step of the proof will be separated out as a lemma.

LEMMA 14. *Let B_1 be an AW^*-algebra of type I. Let B be a sub-C^*-algebra of B_1, and suppose that B is itself an AW^*-algebra.*[6] *Suppose further that B contains all the abelian projections in B_1. Then $B = B_1$.*

Proof. The argument is essentially the same as that used in [3], Lemma 1. We first note that it suffices to prove that B contains an arbitrary projection f of B_1, for B_1 is generated by its projections. Exhibit f as a l. u. b. (in B_1) of abelian projections $\{g_i\}$. By hypothesis the g's are in B, and there they have a possibly different l. u. b. h. It is clear that at least $h \geqq f$. If $h - f$ is non-zero, it contains an abelian projection k, which will lie in B. Since k is orthogonal to f, it is orthogonal to each g_i, and hence to h, a contradiction. Hence $h = f$, as desired.

From the point of view of the structure of AW^*-algebras, the main fact to be recorded is the following:

---

[5] Such an abelian projection always exists in an AW^*-algebra of type I.

[6] We are carefully avoiding any suggestion that B is assumed to be an AW^*-subalgebra in the sense of [3].

COROLLARY. *Let a commutative AW*-algebra A and a cardinal number* $\aleph$ *be given. Then there exists an* $\aleph$*-homogeneous AW*-algebra of type I, with center isomorphic to A, and it is unique up to *-isomorphism.*

11. Derivations. A derivation of an algebra is a linear transformation $a \to a'$ satisfying $(ab)' = a'b + ab'$. The mapping $a \to ax - xa$ is the inner derivation by x. We shall prove:

THEOREM 9. *Every derivation of an AW*-algebra of type I is inner.*

The following lemma (for which I am indebted to I. M. Singer) will be needed.

LEMMA 15. *Every derivation of a commutative C*-algebra with unit is identically* 0.[7]

Proof. It is enough to prove that the derivation vanishes on the general self-adjoint element x. Let M be a maximal ideal. We have to prove $x'(M) = 0$. Write $y = x - x(M)1$. In any derivation of a ring with unit, $1' = 0$; hence $y' = x'$. Write y as a difference of positive elements, say $y = u - v$. Then $u = w^2$ for a suitable element w, and w, like u, will vanish at M. Since $u' = (w^2)' = 2ww'$, we have $u'(M) = 0$. Similarly $v'(M) = 0$, and so $y'(M) = x'(M) = 0$, as desired.

Proof of Theorem 9. Let B be the algebra, and select an abelian projection e which is not annihilated by any non-zero central element. Our first step is to normalize the derivation so as to vanish on e. From $e^2 = e$ we derive $ee' + e'e = e'$. Left and right multiplication by e yields $ee'e = 0$. Set $y = ee' - e'e$. We compute that $ey - ye = ee' + e'e = e'$. Hence if we subtract from the given drivation the inner derivation by y, we get a derivation vanishing on e. Consequently in the rest of the proof we may assume $e' = 0$.

From this we derive $(eb)' = eb'$, so that the derivation induces a linear transformation (say T) on eB. Also $(ebe)' = eb'e$, i. e. there is an induced derivation on eBe. By Lemma 15 this derivation is 0. This implies that T is a module homomorphism of eB, with eB being regarded as an AW^*-module over eBe.

The next step is to prove that T is continuous. By Theorem 6 it suffices

[7] The proof is easily modified to avoid the assumption of a unit element.

to prove that T has an adjoint. The desired adjoint T^* is in fact given simply by $T^* = -T$, for the equation

$$(ebT, ec) = (eb, ecT^*) = -(eb, ecT)$$

reduces to $eb'c^*e + ebc^{*\prime}e = 0$, and this is a consequence of $(ebc^*e)' = 0$.

By Theorem 8, the operator T on eB coincides with right multiplication by a suitable element x of B. We claim that the given derivation coincides with the inner derivation by x, that is, $a' = ax - xa$ for all a in B. Since right multiplication on eB is faithful, it is enough to prove this after an application to the general element eb of eB. That is, we seek to prove

$$eba' = ebax - ebxa. \tag{11}$$

But $ebx = eb'$ and $ebax = e(ba)' = eb'a + ea'b$. On substituting these into the right side of (11) we accomplish the identification with the left side.

12. Automorphisms. In this final section we shall prove the theorem on automorphisms that was mentioned at the beginning of the paper.

THEOREM 10. *Let B be an AW^*-algebra of type I. Then any automorphism of B leaving the center elementwise fixed is inner.*

Let the automorphism be P. We begin as in Theorem 9, selecting an abelian projection e which is not annihilated by any non-zero central element. We examine the element eP, and observe that it is at any rate an idempotent. It is presumably not self-adjoint. This technical obstacle is overcome by a lemma valid in arbitrary C^*-algebras.

LEMMA 16. *Let f be an idempotent in an arbitrary C^*-algebra with unit. Then f is similar to a projection, that is, there exists a regular element p such that $p^{-1}fp$ is self-adjoint.*

Proof. Let $F(t)$ be a continuous real function of the real variable t, satisfying $F(0) = 0$, $F(t) = 1$ for $t \geqq 1$. Define $p = 1 + f - F(f^*f)$. It should no doubt be susceptible to direct verification that this element p does the trick; but we shall prove it by invoking the theory of polynomial identities. We drop down to the closed subalgebra D generated by 1, f and f^*. According to [4], Lemma 5, the C^*-algebra D satisfies the identity that is characteristic of two by two matrices. It follows that every primitive image of D is either one or four-dimensional. Thus it is enough to carry out the computation for the case of two by two matrices. If the element f is either 0 or 1, then $p = 1$,

as it ought to be. Otherwise f is an idempotent of rank one, and we may choose an orthonormal basis such that

$$f = \begin{pmatrix} 1 & 0 \\ x & 0 \end{pmatrix}.$$

The remainder of the computation goes as follows:

$$f^*f = \begin{pmatrix} 1+|x|^2 & 0 \\ 0 & 0 \end{pmatrix}, \quad F(f^*f) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad p = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix},$$

$$2-p = \begin{pmatrix} 1 & 0 \\ -x & 1 \end{pmatrix} = p^{-1}, \quad p^{-1}fp = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

We now complete the proof of Theorem 10. By Lemma 16, an inner automorphism can be applied to the element eP so as to make it self-adjoint. Thus after a change of notation, we can assume that eP itself is self-adjoint. Like e it will be an abelian projection not annihilated by anything in the center. So [3], Lemma 19, applies to tell us that e and eP are equivalent projections. One knows that for finite projections equivalence implies the existence of a unitary element carrying one projection into the other; this is a slight extension of Th. 5. 7 of [2]. By another normalization we may thus assume $eP = e$. Since $eBe = Ze$, Z the center of B ([3], Lemma 10), P leaves eBe elementwise invariant. It follows that P induces a module homomorphism of eB into itself. This module homomorphism is continuous; indeed by [5], Th. 5. 4, P is continuous on all of B.[8] The final steps in the proof go as in Theorem 9. By Theorem 8, the operation of P on eB coincides with right multiplication by an element x. Along with this, the element x^{-1} corresponds to the automorphism P^{-1}. To see that P coincides with the inner automorphism by x it is enough to check $aP = x^{-1}ax$ when applied to $eb \in eB$. But

$$ebx^{-1}ax = e(bP^{-1})ax = e[(bp^{-1})a]P = eb(aP).$$

This concludes the proof of Theorem 10.

Remarks. 1. A slight extension of the argument in Theorem 10 proves the following: any automorphism of an AW^*-algebra of type I can be expressed as the product of an inner automorphism and a *-automorphism. In this version the result may be true for arbitrary C^*-algebras.

---

[8] The continuity of P on eB may also be proved by an explicit construction of its adjoint.

2. If D is a continuous derivation of a Banach algebra, then $e^D = \sum D^n/n!$ is meaningful and is in fact an automorphism. For C^*-algebras it can be seen that e^D leaves the center elementwise fixed. Professor Singer (letter to the author) has proved the following: if e^D is inner, so is D. This shows that from Theorem 10 we can get Theorem 9, restricted to continuous derivations. It is to be observed that in proving Theorem 9 we did not discover the derivation to be continuous till the very end of the proof. At any rate, this focuses attention on the following interesting problem: is every derivation of a C^*-algebra automatically continuous?

UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY

[1] S. Banach, "*Théorie des Opérations Linéaires*," Warsaw, 1932.

[2] I. Kaplansky, "Projections in Banach algebras," *Annals of Mathematics*, vol. 53 (1951), pp. 235-249.

[3] ———, "Algebras of type I," *Annals of Mathematics*, vol. 56 (1952), pp. 460-472.

[4] ———, "Symmetry of Banach algebras," *Proceedings of the American Mathematical Society*, vol. 3 (1952), pp. 397-399.

[5] C. E. Rickart, "The uniqueness of norm problem in Banach algebras," *Annals of Mathematics*, vol. 51 (1950), pp. 615-628.

A NOTE ON LIE k SYSTEM AUTOMORPHISMS.*

By J. K. GOLDHABER.

The object of this note is to generalize the following two theorems:

JACOBSON and RICKART [3]. *Let $\mathfrak{L}$ be a simple Lie ring in which $[\mathfrak{L}, \mathfrak{L}] \neq 0$. Then any Lie triple system homomorphism of $\mathfrak{L}$ onto itself is either a Lie homomorhpism or anti-homomorphism.*

JACOBSON [2]. *Let $\mathfrak{U}$ be a central simple associative algebra of order n^2 over an arbitrary field $\mathfrak{F}$ of characteristic p. Any Lie automorphism of $\mathfrak{U}$ has either the form $A \to T^{-1}AT + (r \cdot \text{trace } A)I$, $r = 0, \cdots, p-1, nr + 1 \neq 0$, or $A \to -T^{-1}A'T + (r \cdot \text{trace } A)I$, $nr - 1 \neq 0$ where $A \to A'$ is an anti-automorhpism of $\mathfrak{U}$ over $\mathfrak{F}$ and I is the identity of $\mathfrak{U}$.*

Let $\mathfrak{L}$ be a Lie ring which admits a field $\mathfrak{F}$ as an operator domain. ψ is said to be a Lie k system automorphism of $\mathfrak{L}$ if ψ is a one to one mapping of $\mathfrak{L}$ onto itself which is linear, i. e.

$$(a_1A_1 + a_2A_2)^\psi = a_1A^\psi{}_1 + a_2A^\psi{}_2$$

for $a_i \in \mathfrak{F}$, $A_i \in \mathfrak{L}$, and is such that

$$[[A_1A_2]A_3] \cdots]A_k]^\psi = [[A^\psi{}_1A^\psi{}_2]A^\psi{}_3] \cdots]A^\psi{}_k]$$

for $A_i \in \mathfrak{L}$.

THEOREM 1. *Let $\mathfrak{L}$ be a simple Lie ring over a field $\mathfrak{F}$ such that $[\mathfrak{L}, \mathfrak{L}] \neq 0$. If ψ is a Lie k system automorphism of $\mathfrak{L}$ then $\psi = \lambda\Phi$ where Φ is a Lie automorphism of $\mathfrak{L}$ and λ is a $(k-1)$st root of unity.*

$\lambda\Phi$ is defined by $A^{\lambda\Phi} = \lambda A^\Phi$ for all $A \in \mathfrak{L}$. We note that if $\lambda = 1$ (-1) then ψ itself is a Lie automorphism (anti-automorphism) of $\mathfrak{L}$.

Let λ_i $(i = 1, 2, \cdots, k-1)$ be a complete set of $(k-1)$st roots of unity and let $\mathfrak{N}_{\lambda_i}$ be the set of all finite sums of elements of the form $[A_1A_2]^\psi - \lambda_i[A^\psi{}_1A^\psi{}_2]$ with $A_i \in \mathfrak{L}$. (If $\mathfrak{F}$ does not contain all the $(k-1)$st roots of unity then we may extend $\mathfrak{F}$ to a field $\mathfrak{F}'$ containing all these roots and ψ may be extended in an obvious manner to a mapping of $\mathfrak{L}$ over $\mathfrak{F}'$.) We shall

---

* Received February 11, 1953; in revised form, August 11, 1953.

show that each $\mathfrak{N}_{\lambda_i}$ is an ideal of $\mathfrak{L}$ and that $[[\mathfrak{N}_{\lambda_1}\mathfrak{N}_{\lambda_2}]\mathfrak{N}_{\lambda_3}]\cdots]\mathfrak{N}_{\lambda_{k-1}}] = 0$. Theorem 1 will then be shown to follow easily.

We show now that $[\mathfrak{N}_{\lambda_i}, \mathfrak{L}] \leqq \mathfrak{N}_{\lambda_i}$. Since $\mathfrak{L}$ is simple and $[\mathfrak{L}, \mathfrak{L}] \neq 0$, and since $\mathfrak{L}^\psi = \mathfrak{L}$ it follows that if $L \in \mathfrak{L}$ then L is a sum of elements of the form $[[L^\psi_1 L^\psi_2]L^\psi_3]\cdots]L^\psi_{k-1}]$, with $L_i \in \mathfrak{L}$. But now,

$$
\begin{aligned}
&[[A_1A_2]^\psi - \lambda_i[A^\psi_1 A^\psi_2], [[L^\psi_1 L^\psi_2]\cdots]L^\psi_{k-1}]] \\
&\quad = [[A_1A_2]^\psi[[L^\psi_1 L^\psi_2]\cdots]L^\psi_{k-1}]] - \lambda_i[[A^\psi_1 A^\psi_2][[L^\psi_1 L^\psi_2]\cdots]L^\psi_{k-1}]] \\
&\quad = [[A_1A_2][[L_1L_2]\cdots]L_{k-1}]]^\psi - \lambda_i[[[[L^\psi_1 L^\psi_2]\cdots]L^\psi_{k-1}]A^\psi_2], A^\psi_1] \\
&\qquad\qquad - \lambda_i[[A^\psi_1[[L^\psi_1 L^\psi_2]\cdots]L^\psi_{k-1}]], A^\psi_2] \\
&\quad = [[[[L_1L_2]\cdots]L_{k-1}]A_2], A_1]^\psi - \lambda_i[[[[L_1L_2]\cdots]L_{k-1}]A_2]^\psi, A^\psi_1] \\
&+ [[A_1[[L_1L_2]\cdots]L_{k-1}]], A_2]^\psi - \lambda_i[[A_1[[L_1L_2]\cdots]L_{k-1}]]^\psi, A^\psi_2] \in \mathfrak{N}_{\lambda_i}.
\end{aligned}
$$

Hence $[[A_1A_2]^\psi - \lambda_i[A^\psi_1 A^\psi_2], L] \in \mathfrak{N}_{\lambda_i}$. In the above calculations use has been made of the Jacobi identity $[[AB]C] + [[BC]A] + [[CA]B] = 0$. It has thus been shown that $\mathfrak{N}_{\lambda_i}$ is an ideal of $\mathfrak{L}$.

We now show that $[\mathfrak{N}_{\lambda_1}, \mathfrak{N}_{\lambda_2}]$ consists of all finite sums of the form

$$
\begin{aligned}
[[A_{11}A_{21}]^\psi[A_{12}A_{22}]^\psi] &- (\lambda_1 + \lambda_2)[[A_{11}A_{21}]^\psi[A^\psi_{12}A^\psi_{22}]] \\
&+ \lambda_1\lambda_2[[A^\psi_{11}A^\psi_{21}][A^\psi_{12}A^\psi_{22}]].
\end{aligned}
$$

For

$$
\begin{aligned}
&[[A_{11}A_{21}]^\psi - \lambda_1[A^\psi_{11}A^\psi_{21}], [A_{12}A_{22}]^\psi - \lambda_2[A^\psi_{12}A^\psi_{22}]] \\
&\quad = [[A_{11}A_{21}]^\psi[A_{12}A_{22}]^\psi] - \lambda_1[[A^\psi_{11}A^\psi_{21}][A_{12}A_{22}]^\psi] \\
&\qquad - \lambda_2[[A_{11}A_{21}]^\psi[A^\psi_{12}A^\psi_{22}]] + \lambda_1\lambda_2[[A^\psi_{11}A^\psi_{21}][A^\psi_{12}A^\psi_{22}]].
\end{aligned}
$$

The desired result will be established as soon as it is shown that

$$[[A_{11}A_{21}]^\psi[A^\psi_{12}A^\psi_{22}]] - [[A^\psi_{11}A^\psi_{21}][A_{12}A_{22}]^\psi] = B = 0.$$

As above, if $L \in \mathfrak{L}$ then there exist $L_i \in \mathfrak{L}$ such that L is a sum of elements of the form $[[L^\psi_1 L^\psi_2]L^\psi_3]\cdots]L^\psi_{k-3}] = M$. But then if we expand $[B, M]$ and use the fact that ψ preserves Lie k products we find that it is equal to zero. The desired equality then follows from the simplicity of $\mathfrak{L}$.

An induction argument may now be used to show that

$$[[\mathfrak{N}_{\lambda_1}, \mathfrak{N}_{\lambda_2}]\mathfrak{N}_{\lambda_3}]\cdots]\mathfrak{N}_{\lambda_{k-1}}]$$

consists of all finite sums of elements of the form

$$[[[A_{11}A_{21}]^{\psi}[A_{12}A_{22}]^{\psi}][A_{13}A_{23}]^{\psi}]\cdots][A_{1k-1}A_{2k-1}]^{\psi}]$$
$$-(\sum_i \lambda_i)[[[A_{11}A_{21}]^{\psi}[A_{12}A_{22}]^{\psi}]\cdots][A_{1k-2}A_{2k-2}]^{\psi}][A^{\psi}_{1k-1}A^{\psi}_{2k-1}]]$$
$$+(\sum_{i<j}\lambda_i\lambda_j)[[[A_{11}A_{21}]^{\psi}[A_{12}A_{22}]^{\psi}]\cdots][A_{1k-3}A_{2k-3}]^{\psi}][A^{\psi}_{1k-2}A^{\psi}_{2k-2}]][A^{\psi}_{1k-1}A^{\psi}_{2k-1}]]$$
$$-\cdots\cdots\cdots+(-1)^{k-1}\prod_i \lambda_i[[[A^{\psi}_{11}A^{\psi}_{21}][A^{\psi}_{12}A^{\psi}_{22}]\cdots][A^{\psi}_{1k-1}A^{\psi}_{2k-1}]].$$

Since the λ_i form a complete set of $(k-1)$st roots of unity it is true that all the elementary symmetric functions of the λ_i except $\prod\lambda_i$ are zero, the latter being equal to $(-1)^k$. Furthermore we may show in a manner similar to the one used above that

$$[[[A_{11}A_{21}]^{\psi}[A_{12}A_{22}]^{\psi}]\cdots][A_{1k-1}A_{2k-1}]^{\psi}]$$
$$=[[[A^{\psi}_{11}A^{\psi}_{21}][A^{\psi}_{12}A^{\psi}_{22}]]\cdots][A^{\psi}_{1k-1}A^{\psi}_{2k-1}]].$$

But then it follows that $[[\mathfrak{N}_{\lambda_1},\mathfrak{N}_{\lambda_2}]\mathfrak{N}_{\lambda_3}]\cdots]\mathfrak{N}_{\lambda_{k-1}}]=0$. Now since the $\mathfrak{N}_{\lambda_i}$ are ideals of $\mathfrak{L}$ and since $\mathfrak{L}$ is simple it follows that one of the $\mathfrak{N}_{\lambda_i}=0$. But then we have that for some $(k-1)$st root of unity, λ, and for all $A_i\,\varepsilon\,\mathfrak{L}$ $[A_1A_2]^{\psi}=\lambda[A^{\psi}_1A^{\psi}_2]$. (We note parenthetically that this indicates that $\mathfrak{L}$ over $\mathfrak{F}$ must admit λ as a multiplicative operator; i. e. we may consider λ as belonging to $\mathfrak{F}$.) Then $[A_1A_2]^{\lambda\psi}=\lambda[A_1A_2]^{\psi}=\lambda^2[A^{\psi}_1A^{\psi}_2]=[A^{\lambda\psi}_1A^{\lambda\psi}_2]$ and $\lambda\psi=\Phi$ where Φ is a Lie automorphism of $\mathfrak{L}$. Hence $\psi=\lambda^{-1}\Phi$ as desired.

It may be of interest to state Theorem 1 in the following equivalent form:

THEOREM 1a. *Let $\mathfrak{L}$ be as in Theorem 1. Let $\mathfrak{G}_k$ be the group of all Lie k system automorphisms of $\mathfrak{L}$ and let $\mathfrak{H}$ be the group of all Lie automorphisms of $\mathfrak{L}$. Then $\mathfrak{H}$ is an invariant subgroup of $\mathfrak{G}_k$ and $\mathfrak{G}_k/\mathfrak{H}$ is isomorphic with the multiplicative group of the $(k-1)$st roots of unity which are contained in the ground field $\mathfrak{F}$.*

We now generalize Jacobson's theorem concerning the structure of Lie automorphisms of simple algebras to Lie k system automorphisms of semi-simple separable algebras.

THEOREM 2. *Let $\mathfrak{S}$ be a semi-simple separable algebra over an arbitrary field $\mathfrak{F}$ of characteristic p. To every simple component $\mathfrak{U}$ of $\mathfrak{S}$ there corresponds a simple component $\mathfrak{B}$, isomorphic with $\mathfrak{U}$, such that any Lie k system automorhpism of $\mathfrak{S}$ has on $\mathfrak{U}$ either of the following two forms:*

I. $A\to\lambda[B+(r\,\mathrm{trace}\,B)I]+(\mathrm{trace}\,B)Z, \qquad r=0,1,\cdots,p-1$

II. $A\to\lambda[-B'+(r\,\mathrm{trace}\,B')I]+(\mathrm{trace}\,B')Z \qquad r=0,1,\cdots,p-1$

where $A \to B(B')$ is an isomorphism (anti-isomorphism) of $\mathfrak{U}$ and $\mathfrak{B}$, λ a $(k-1)$st root of unity, and Z an element of the center of $\mathfrak{S} - \mathfrak{B}$.

Let $\mathfrak{K}$ be a splitting field for $\mathfrak{S}$. Then $\mathfrak{S}_{\mathfrak{K}} = \mathfrak{M}_1 + \mathfrak{M}_2 + \cdots + \mathfrak{M}_m$ where the $\mathfrak{M}_i$ are total matric algebras over $\mathfrak{K}$. It will be sufficient to determine the form of a Lie k system automorphism, ψ, on each $\mathfrak{M}_i$.

Let V_1 be a non-zero linear subset of $\mathfrak{M}' = [\mathfrak{M}, \mathfrak{M}]$ where $\mathfrak{M}$ is a total matric algebra. Let $V_j = [[\overbrace{[\mathfrak{M}, \mathfrak{M}] \cdots]\mathfrak{M}}^{k-1}]V_{j-1}]$. If the degree of $\mathfrak{M}$ is not equal to zero in the base field then $\mathfrak{M}'$ is simple [2, Theorem 7] and it follows that there exists a positive integer q such that $V_q = M'$. If the degree of $\mathfrak{M}$ equals zero a computational argument may be used to show that the same result holds provided it is assumed that V_1 contains an element not in the center of $\mathfrak{M}$.

Henceforth we use the notation $\psi(A)$ instead of A^{ψ}.

Suppose now that for some $M \in \mathfrak{M}_i$ $\psi(M) = M_1 + M_2 + \cdots + M_m$ where $M_j \in \mathfrak{M}_j$ and where M_h (say) does not belong to the center of $\mathfrak{M}_h$. Then there exist $M_{hj} \in \mathfrak{M}_h$ $(j = 1, 2, \cdots, k-1)$ such that

$$[[[M_h, M_{h1}]M_{h2}] \cdots]M_{hk-1}] \in \mathfrak{M}'_h$$

(but not in the center of $\mathfrak{M}_h$). Since ψ is one to one there exist $S_j \in \mathfrak{S}$ such that $\psi(S_j) = M_{hj}$. But then

$$M' = [[[M, S_1]S_2] \cdots]S_{k-1}] \in \mathfrak{M}'_i$$

and

$$\begin{aligned}\psi(M') &= [[[\psi(M), \psi(S_1)]\psi(S_2)] \cdots]\psi(S_{k-1})] \\ &= [[[M_h, M_{h1}]M_{h2}] \cdots]M_{hk-1}] \in \mathfrak{M}'_h.\end{aligned}$$

Hence if there exists an $M \in \mathfrak{M}_i$ such that $\psi(M)$ has a non-central projection in $\mathfrak{M}_h$ then there exists an $M' \in \mathfrak{M}'_i$ such that $\psi(M') \in \mathfrak{M}'_h$, $\psi(M')$ not in the center of $\mathfrak{M}_h$.

Let V_1 be the set spanned by M'. If V_j is defined as above then $\psi(V_j) \leqq \mathfrak{M}'_h$ for $j = 1, 2, \cdots$. It follows that $\psi(\mathfrak{M}'_i) \leqq \mathfrak{M}'_h$. But we also have $\psi(\mathfrak{M}'_i) = \mathfrak{M}'_h$; for now let V_1 be the set of all elements in $\mathfrak{M}'_h$ which are not the maps of any element in $\mathfrak{M}'_i$. Suppose that V_1 is not empty. Then V_1 has a basis such that the inverse image, ψ^{-1}, of any element in this basis has 0 as its M_i-th component. Since $V_q = \mathfrak{M}'_h$ it follows that every element of $\mathfrak{M}'_h$ is the map of an element of $\mathfrak{S}$ whose $\mathfrak{M}_i$-th component is zero. In view of the facts that ψ is one to one and that $\psi(\mathfrak{M}'_i) \leqq \mathfrak{M}'_h$ this is impossible. Hence V_1 is empty and $\psi(\mathfrak{M}'_i) = \mathfrak{M}'_h$. (This argument

holds only for the case that $\mathfrak{M}'_h$ is Lie simple. A different argument must be used in the case that the degree of $\mathfrak{M}_h$ is 0 in the base field.) We note furthermore that this implies that $\mathfrak{M}_i$ and $\mathfrak{M}_h$ have the same order and hence they are isomorphic.

Now let E_{st} be a basis for $\mathfrak{M}_i$. From the above it follows that $\psi(E_{ss})$ cannot have a non-central projection outside of $\mathfrak{M}_h$. Furthermore since $E_{ss} - E_{tt} \in \mathfrak{M}'_i$ it also follows that the part of $\psi(E_{ss})$ which lies outside of $\mathfrak{M}_h$ is the same for all s. Hence $\psi(E_{ss}) = J_s + Z$ where $J_s \neq 0 \in \mathfrak{M}_h$ and Z is in the center of $\mathfrak{S} - \mathfrak{M}_h$. We have thus shown that for all $M \in \mathfrak{M}_i$ $\psi(M) = \Phi(M) + (\text{trace } M)Z$ where Φ is a Lie k system isomorphism of $\mathfrak{M}'_i$ onto $\mathfrak{M}'_h$. Since $\mathfrak{M}_i$ and $\mathfrak{M}_h$ are isomorphic the theorem now follows easily from Theorem 1 and Jacobson's theorem.

COROLLARY. *Let $\mathfrak{S}$ be a semi-simple separable algebra over $\mathfrak{F}$ such that the principal degrees of the simple components of $\mathfrak{S}$ are all distinct and all different from zero in $\mathfrak{F}$. If ψ is a Lie automorphism of $\mathfrak{S}$ which leaves the elements of the center invariant then ψ is an automorphism of $\mathfrak{S}$.*

CORNELL UNIVERSITY.

REFERENCES.

[1] N. Jacobson, "Simple Lie algebras over a field of characteristic zero," *Duke Mathematical Journal*, vol. 4 (1938), pp. 534-551.

[2] ———, "Classes of restricted Lie algebras of characteristic p, I," *American Journal of Mathematics*, vol. 63 (1941), pp. 481-515.

[3] ——— and C. E. Rickart, "Jordan homomorphism of rings," *Transactions of the American Mathematical Society*, vol. 69 (1950), pp. 479-502.

THE STRUCTURE OF A CERTAIN CLASS OF RINGS.*

By I. N. Herstein.

In [2], [3] and [4] we considered certain fairly general assumptions which, when imposed on a given ring, rendered it commutative. In this paper we carry this type of investigation further and extend the previously obtained results.

Let R be a ring with center Z. For $a \in R$, $p_a(t)$ will denote a polynomial in the indeterminate t having rational integers as coefficients, where we further suppose that these coefficients are functions of a. Then $p_a(a)$ will denote that ring element in R which is obtained upon substituting a for t in the formal polynomial $p_a(t)$.

In this paper we consider rings with the property that for every $a \in R$ there exists such a polynomial $p_a(t)$ so that $a^2 p_a(a) - a$ is in Z. For this class of rings we prove the

Theorem. *If R is a ring with center Z such that for every $a \in R$ there exists a polynomial $p_a(t)$ such that $a^2 p_a(a) - a \in Z$, then R is commutative.*

This result subsumes that of [4], for there $p_a(t) = t^{m(a)}$. For polynomials having linear terms it is a generalization of the theorem, using a fixed polynomial, proved for semi-simple rings by Kaplansky [7]. As Kaplansky pointed out by an example, generalizations which do not involve polynomials with linear terms would probably require fairly restrictive hpotheses.

1. A field theoretic theorem. We begin with a theorem concerning fields. This will be applied to study the division ring case of our theorem.

Theorem 1. *Let $K \supset Z$, $K \neq Z$ be two fields with K being a finite extension of Z. Suppose that for every $a \in K$ there exists a polynomial $p_a(t)$ with integral coefficients so that $a^{r+1} p_a(a) - a^r \in Z$ ($r > 0$ depending on a). Then Z is of characteristic $p \neq 0$ and either (1) K is purely inseparable over Z or (2) Z, and so K, is algebraic over its prime field.*

---

* Received April 13, 1953.

Proof. Let us suppose that K is not purely inseparable over Z. If K is of characteristic 0 or if Z is of characteristic $p \neq 0$ but is not algebraic over its prime field, then by a theorem of Nakayama [8] there exist two (non-archimedean) valuations V_1 and V_2 which differ on K but coincide on Z. Thus there is an $x \in K$ such that $V_1(x) \neq V_2(x)$. Without loss of generality we may assume that $V_1(x) > 0$, for otherwise we would use x^{-1} instead of x. Since these two valuations coincide on Z, they must induce the same valuation on P, the prime field of Z. If P is finite, then $V_i(p) \geqq 0$ for all $p \in P$, since V_i is then a trivial valuation on P. If P is the field of rational numbers, then V_i induces some p-adic valuation on P, so $V_i(n) \geqq 0$ for all integers n. Now $\alpha = x^{r+1}p(x) - x^r \in Z$, where $p(x)$ is a polynomial with integral coefficients; thus $V_1(\alpha) = V_2(\alpha)$. Since $V_1(x) > 0$ and $V_1(n_i) \geqq 0$ for n_i an integer, $rV_1(x) = V_1(x^r) < V_1(n_i x^{r+i}) = V_1(n_i) + (r+i)V_1(x)$. Thus $V_1(\alpha) = V_1(x^r) = rV_1(x)$. Similarly $V_2(\alpha) = rV_2(x)$. Then $V_1(\alpha) = V_2(\alpha)$ forces $V_1(x) = V_2(x)$, a contradiction. So Z must be both of characteristic $p \neq 0$ and algebraic over its prime field. This completes the proof of Theorem 1.

In a certain sense this theorem is a generalization of a result due to Ikeda [5]. He assumes that the polynomial $p_a(t)$ have fixed integral coefficients, but, on the other hand, does not require that the term of lowest degree should have coefficient ± 1.

2. The semi-simple case. We apply Theorem 1 to prove

THEOREM 2. *If D is a division ring with center Z such that $a^{r+1}p_a(a) - a^r \in Z$ ($r > 0$ depending on a) for every $a \in R$, then D is commutative.*

Proof. If D is not commutative, by a theorem of Noether (as generalized by Jacobson [6]), there exists an element $a \notin Z$, $a \in D$, which is separable over Z. Consider $K = Z(a)$. The conditions of Theorem 1 hold true for the two fields $K \supset Z$, and since K is not purely inseparable over Z, Z must be of characteristic $p \neq 0$ and algebraic over its prime field. D is thus algebraic over its prime field, which is a finite field. This possibility is ruled out by a theorem of Jacobson [6] which states that there are no non-commutative algebraic division algebras over a finite field. D is thus a commutative field.

For general rings we cannot get by with the "liberal" conditions on Theorem 2; we must assume that $r = 1$. All the rings R henceforth considered in this paper will be such that $a^2p_a(a) - a \in Z$, $p_a(t)$ a polynomial with integral coefficients, for every a in R.

LEMMA 3. *If $a \in R$ is nilpotent then $a \in Z$.*

For suppose $a^{2^n} = 0$. Since $a^2 p_1(a) - a \in Z$,

$$[a^2 p_1(a)]^2 q(a^2 p_1(a)) - a^2 p_1(a) = a^4 p_2(a) - a^2 p_1(a) \in Z, \cdots$$

$a^{2^n} p_n(a) - a^{2^{n-1}} p_{n-1}(a) \in Z$, we obtain, by addition, $a^{2^n} p_n(a) - a \in Z$. Since $a^{2^n} = 0$, it follows that $a \in Z$.

An immediate consequence of Lemma 3 is

LEMMA 4. *If $e \in R$ and $e^2 = e$, then $e \in Z$.*

For $(xe - exe)^2 = (ex - exe)^2 = 0$ for all $x \in R$. By Lemma 3 both $xe - exe$ and $ex - exe$ are then in Z. So $0 = e(xe - exe) = (xe - exe)e = xe - exe$, and likewise $ex - exe = 0$, hence $xe = ex$.

Using these two lemmas we are able to establish

THEOREM 5. *If R is primitive then it is a commutative field.*

Proof. Since R is a primitive ring it possesses a maximal right ideal ρ which contains no non-zero ideal of R (*by ideal we mean here and henceforth a two-sided ideal*). Let $a \in \rho$. Thus $b = a^2 p_a(a) - a \in \rho \cap Z$. Since $b \in Z$, $bR \subset \rho$ is an ideal of R, so $bR = (0)$. Since R is primitive this is only possible if $b = 0$, that is, if $a^2 p_a(a) = a$. But then $e = a p_a(a)$ is an idempotent and is in ρ; moreover by Lemma 4 e is in Z. Thus $eR \subset \rho$ is an ideal of R, so again $eR = (0)$. This forces $a = ae = 0$. That is, (0) is a maximal right ideal of R. Since R is primitive, it must be a division ring, and so it is a commutative field by Theorem 2.

As an immediate corollary we have the corollary: *If R is semi-simple, then it is commutative.*

3. The general case. A ring is said to be subdirectly irreducible if the intersection of its non-zero ideals is a non zero-ideal. Every ring is isomorphic to a subdirect sum of subdirectly irreducible rings [1]. Since the property $a^2 p_a(a) - a \in Z$ is preserved under a homomorphism, it is enough, in order to settle the general case, to establish it for the particular case in which R is a subdirectly irreducible ring.

We henceforth assume that R is a subdirectly irreducible ring with $S \neq (0)$ the intersection of its non-zero ideals. Our purpose is to show that R is commutative. If R should be semi-simple, we know this to be so by the corollary of Theorem 5. We may then assume that the Jacobson radical, J, of R is not the zero ideal.

Since $J \neq (0)$ and is an ideal of R, by its very definition $S \subset J$, where S is the minimal ideal of R. As a ring by itself, S can be one of two types: either S is a trivial ring, that is, $S^2 = (0)$, or S is a simple ring. In the latter case, since S is a simple radical ring, if $S^2 \neq (0)$, then S can have only a trivial center, $Z = (0)$. This would lead to $a^2 p_a(a) = a$ for every a belonging to S. Then S would be a regular ring, and would therefore be semi-simple, contradicting the fact that S is a radical ring. The second alternative is, in this way, one which cannot actually occur, and we are left with $S^2 = (0)$. All in all we have proved

THEOREM 6. $S^2 = (0)$.

This, in conjunction with Lemma 3, immediately gives us

THEOREM 7. $S \subset Z$.

Let $A(S) = \{x \in R \mid Sx = (0)\}$. Then $A(S)$ is an ideal of R, and since $S \subset A(S)$ by Theorem 6, $A(S)$ is not the zero ideal.

Suppose $x, y \in R$. For any $s \in S$ $sx \in S \subset Z$, so $(sx)y = y(sx) = syx$, since $s \in Z$. That is, $s(xy - yx) = 0$. Thus we have proved

THEOREM 8. *For all* $x, y \in R$, $(xy - yx) \in A(S)$.

LEMMA 9. *If* $e \in R$ *and* $e^2 = e$, *then either* $e = 0$ *or* $e = 1$ *if the ring possesses a unit element.*

Proof. Since $e^2 = e$, by Lemma 4 $e \in Z$. Thus Re is an ideal of R. Similarly, $V = \{x \in R \mid x = re - r, r \in R\}$ is also an ideal of R. But $S \subset V \cap Re = (0)$ if $V \neq (0)$ and $Re \neq (0)$, a contradiction. So either $V = (0)$ or $Re = (0)$, proving the lemma.

LEMMA 10. *If* ρ *is a non-zero right ideal of* R, *then* $S \subset \rho$. *Likewise, if* λ *is a non-zero left-ideal of* R, $S \subset \lambda$.

Proof. Let $x \neq 0 \in \rho$. Then $y = x^2 p_x(x) - x \in \rho \cap Z$. If $y = 0$ then $x^2 p_x(x) = x \neq 0$, which leads to $e = x p_x(x)$ a non-zero idempotent; but then $x p_x(x) = 1$ by Lemma 9, so $1 \in \rho$, and $\rho = R$ in which case vacuously $S \subset \rho$. If $Ry \neq (0)$, then it is an ideal of R and is contained in ρ, so $S \subset Ry \subset \rho$. We have only one situation left to consider, namely $Ry = 0$, $y \neq 0$. But then $T = \{b \in R \mid Rb = (0)\}$ is a non-zero ideal of R; a simple check yields that $T \cap \rho \subset \rho$ is also a non-zero ideal of R (non-zero, since $y \in T \cap \rho$).

Hence $S \subset T \cap \rho \subset \rho$. Clearly, an analogous argument establishes the result in case of a non-zero left ideal.

A key result for the later results of this paper is

THEOREM 11. *If* $a \in A(S)$, *then* $a \in Z$.

Proof. Suppose that $a \in A(S)$ and $a \notin Z$. For some $y \in R$, $v = ay - ya \neq 0$. If $Rv \neq (0)$, then $S \subset Rv$, by Lemma 10. So if $s \neq 0 \in S$, then $s = rv = r(ay - ya)$ for some $r \in R$. Let $b = v^2 p_v(v) - v \in Z$. Then $b \neq 0$, for if $b = 0$, then $vp_v(v)$ is a non-zero idempotent, hence 1, by Lemma 9; this situation is impossible, since $v \in A(S)$ is a zero-divisor, by Theorem 8. We may thus assume that $b \neq 0$. Since $a \in A(S)$, $0 = sa = rva$. But knowing that $v \in A(S)$ we also have that $rba = -rva$, so $rba = 0$. However, $b \in Z$, so $0 = rba = rab$. That is, $0 = rav^2 p_v(v) - rav$ or, equivalently, $(rav)(vp_v(v)) = rav$.

Let $V = \{x \in R \mid xvp_v(v) = x\}$. Then V is a left-ideal of R, so if $V \neq (0)$ we would have that $S \subset V$ by Lemma 10. That is, if $s \neq 0 \in S$, $svp_v(v) = s$; this, together with $v \in A(S)$, would yield $0 \neq s = svp_v(v) = 0$, a contradiction. We are forced to conclude that $V = (0)$. As a consequence, $0 = rav = ra(ay - ya)$, since $rav \in V$. Similarly $ra^n(ay - ya) = 0$ for all $n \geqq 1$. Now $r(a^2y - ya^2) = r[a(ay - ya) + (ay - ya)a] = 0 + sa = 0$, since $a \in A(S)$. Since $r(a^iy - ya^i) = r[a^{i-1}(ay - ya) + (a^{i-1}y - ya^{i-1})a]$, an obvious induction leads to $r(a^iy - ya^i) = 0$ for all $i \geqq 2$. Since

$$a^2 p_a(a) - a = \sum_{i=2}^{n} \alpha_i a^i - a \in Z,$$

where the α_i are rational integers,

$$ay - ya = \sum_{i=2}^{n} \alpha_i (a^iy - ya^i)$$

and so

$$0 \neq s = r(ay - ya) = \sum_{i=2}^{n} \alpha_i r(a^iy - ya^i) = 0$$

by the above remarks. Since this is impossible, we conclude that $Rv = (0)$. But then $v^2 = (ay - ya)^2 = 0$, so $ay - ya \in Z$, by Lemma 3. Since $R(ay - ya) = (ay - ya)R = (0)$, $a(ay - ya) = 0 = (ay - ya)a$. Thus $a^2y = ya^2$. Continuing in this way we obtain $a^iy = ya^i$ for all $i \geqq 2$. Thus

$$ay - ya = \sum_{i=2}^{n} \alpha_i (a^iy - ya^i) = 0,$$

contradicting $ay - ya \neq 0$. Theorem 11 is thereby established.

Although it is but a special case of Theorem 11, for future use we single out

THEOREM 12. *For all* $x, y \in R$, $xy - yx \in Z$.

This is immediate from the fact that $xy - yx \in A(S)$ by Theorem 8 together with Theorem 11. Theorem 11 also yields

THEOREM 13. *If* $R = A(S)$, *then* R *is commutative.*

From now on we assume that $A(S) \neq R$.

LEMMA 14. *If* $s \neq 0 \in S$, *then* $Rs = S$.

For if $s \in S$, then $s \in Z$, so Rs is an ideal of R and $Rs \subset S$. If $Rs = (0)$, then $T = \{x \in R \mid Rx = (0)\}$ is a non-zero ideal of R, so $S \subset T$, and $RS = (0) = SR$. This implies that $A(S) = R$, a situation we have already settled and ruled out. Thus $Rs \neq (0)$, and so $S \subset Rs \subset S$.

We can now prove

THEOREM 15. $R/A(S)$ *is a field.*

Proof. Suppose that $x \in R$, $x \notin A(S)$.

If $s \neq 0 \in S$, then $sx \neq 0$, for $sx = 0$ implies $Rsx = (0)$, and so, by Lemma 14, $Sx = (0)$, implying the false result $x \in A(S)$. Since $sx \neq 0 \in S$, using Lemma 14 again, $Rsx = S$. For some $y \in R$ we have $s = ysx = syx$, since $s \in Z$. Let $e = yx$. For all $r \in R$, $s(re - r) = 0$, so $Rs(re - r) = (0)$, and so $S(re - r) = 0$, hence $re - r \in A(S)$. If $u \notin A(S)$, the same argument used in exhibiting e leads to the existence of a $w \in R$ with $uw - e \in A(S)$. In this way $R/A(S)$ is a division ring with $e + A(S)$ as its unit element. Since all $xy - yx \in A(S)$, $R/A(S)$ is commutative. All told, $R/A(S)$ is a field.

Let $a \in R$, $a \notin A(S)$. Suppose that $q(t)$ is a polynomial of lowest positive degree having rational integer coefficients and that $q(a) \in Z$. If $a \notin Z$, then $ay - ya \neq 0$ for some $y \in R$. Now $q(a)y - yq(a) = 0$, and since $ay - ya \in Z$ by Theorem 12, $0 = q(a)y - yq(a) = q'(a)(ay - ya)$, where $q'(t)$ is the formal derivative of the polynomial $q(t)$. Thus $q'(a)$ is a zero-divisor; from this we conclude (as we have done in several proofs before) that $q'(a)$ must be in $A(S)$. Since $A(S) \subset Z$ by Theorem 11, $q'(a) \in Z$. But since $q(t)$ was the polynomial of least *positive* degree with $q(a) \in Z$, it follows that $q'(a) =$ an element of Z. Since

$$q(a) = \alpha_0 + \alpha_1 a + \cdots + \alpha_i a^i + \cdots + \alpha_n a^n,$$

where the α_i are rational integers,

$$q'(a) = \alpha_1 + \cdots + i\alpha_i a^{i-1} + \cdots + n\alpha_n a^{n-1}.$$

Thus $i\alpha_i a^{i-1} = 0$ for all $i > 1$. Now $\alpha_i \neq 0$ for some $i > 1$, since $q(t)$ has positive degree, thus in $R/A(S)$ $i\alpha_i(\bar{a})^{i-1} = 0$ with $(\bar{a})^{i-1} \neq 0$. Consequently, $R/A(S)$ is of characteristic $p \neq 0$. We have, in this discussion, proved

THEOREM 16. *$R/A(S)$ is of characteristic $p \neq 0$.*

Let P be the prime field of $\bar{R} = R/A(S)$. Then P has p elements. Now if $a \notin A(S)$ and $a \notin Z$, then $ay - ya \neq 0$ for some $y \in R$. Since $a^2 p_a(a) - a = \sum_{i=2}^{n} \alpha_i a^i - a \in Z$, where the α_i are rational integers

$$(*) \qquad (\sum_{i=2}^{n} \alpha_i a^i)y - y(\sum_{i=2}^{n} \alpha_i a^i) = ay - ya.$$

However, $ay - ya \in Z$, so the left side becomes

$$(\sum_{i=2}^{n} i\alpha_i a^{i-1})(ay - ya)\text{; hence}$$

$$[(\sum_{i=2}^{n} i\alpha_i a^i) - a](ay - ya) = 0,$$

by (*). Since $ay - ya \neq 0$, $(\sum_{i=2}^{n} i\alpha_i a^i) - a$ is a zero divisor, so must be in $A(S)$. In $R/A(S)$ this leads to the equation $\sum_{i=2}^{n} i\alpha_i \bar{a}^i - \bar{a} = 0$, so $\bar{a}$ is algebraic over the prime field P. $P(\bar{a})$ is thus a finite field. This, of course, implies that for some integer $n(\bar{a}) > 1$, $\bar{a}^{n(\bar{a})} = \bar{a}$. In R this becomes $a^{n(\bar{a})} - a \in A(S) \subset Z$. If $a \notin A(S)$ and $a \in Z$, then obviously $a^{n(a)} - a \in Z$. Likewise, if $a \in A(S)$, then $a \in Z$, so $a^{n(a)} - a \in Z$. In other words, for all $x \in R$, $x^{n(x)} - x \in Z$ for some $n(x) > 1$. By the main result of [4], R must be commutative. We have thus proved Theorem 17. *If $A(S) \neq R$ then R is commutative.*

Between Theorems 13 and 17 we have taken care of all possibilities when R is subdirectly irreducible. So we have

THEOREM 18. *If R is subdirectly irreducible then it is commutative.*

Using the decomposition of a general ring as a subdirect sum of

subdirectly irreducible ones we have completed the proof of the main theorem of the paper, namely

THEOREM 19. *If in R every element a satisfies a relation of the form $a^2 p_a(a) - a \in Z$ then R is commutative.*

UNIVERSITY OF CHICAGO
AND
UNIVERSITY OF PENNSYLVANIA.

REFERENCES.

[1] G. Birkhoff, "Subdirect unions in universal algebra," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 764-768.

[2] I. N. Herstein, "A generalization of a theorem of Jacobson," *American Journal of Mathematics*, vol. 73 (1951), pp. 756-762.

[3] ———, "A generalization of a theorem of Jacobson II," Abstract, *Bulletin of the American Mathematical Society*, vol. 58 (1952), p. 474.

[4] ———, "A generalization of a theorem of Jacobson III," *American Journal of Mathematics*, vol. 75 (1953), pp. 105-111.

[5] M. Ikeda, "On a theorem of Kaplansky," *Osaka Mathematical Journal*, vol. 4 (1952), pp. 235-240.

[6] N. Jacobson, "Structure theory for algebraic algebras of bounded degree," *Annals of Mathematics*, vol. 46 (1947), pp. 695-707.

[7] I. Kaplansky, "A theorem on division rings," *Canadian Journal of Mathematics*, vol. 3 (1951), pp. 290-292.

[8] T. Nakayama, "The commutativity of division rings," *Canadian Journal of Mathematics*, vol. 5 (1953), pp. 242-244.

ERRATA.

V. L. Shapiro, "Square summation and localization of double trigonometric series," this JOURNAL, vol. 75, 347-357.

Page 348, in Lemma 2, condition (i), read $\sum_{|M|\geqq R}$ instead of $\sum_{|M|\leqq R}$.

Page 348, in Lemma 2, condition (ii), read $\sum_{|M|\leqq R}$ instead of $\sum_{|M|\geqq R}$.

Page 349, line 3 from below, read $\gamma, \eta > 0$, instead of $\gamma, \delta\eta \geqq 0$.

Page 350, line 5 from above, read $\gamma, \eta \geqq 0$ instead of $\delta, \eta \geqq 0$.

Page 353, line 13 from above, read $\delta^k T_2/\delta y^k$ instead of $\delta^k T_2/\delta^k$.

American Journal of Hygiene. Edited by D. BODIAN, Managing Editor, A. M. BAETJER, R. M. HERRIOTT, K. F. MAXCY, M. MERRELL, and T. B. TURNER. Publishing two volumes of three numbers each year, volume 58 is now in progress. Subscription $12 per year. (Foreign postage, 50 cents; Canadian postage, 25 cents.)

American Journal of Mathematics. Edited by R. BAER, WEI-LIANG CHOW, S. EILENBERG and A. WINTNER. Quarterly. 8vo. Volume 75 in progress. $8.50 per volume. (Foreign postage, 50 cents; Canadian, 25 cents.)

American Journal of Philology. Edited by H. T. ROWELL, LUDWIG EDELSTEIN, K. MALONE, B. D. MERITT, JAMES H. OLIVER, and EVELYN H. CLIFT, Secretary. Volume 74 is in progress. $6.00 per volume. (Foreign postage, 50 cents; Canadian, 25 cents.)

Bulletin of the History of Medicine. OWSEI TEMKIN, Editor. Bi-monthly. Volume 27 in progress. 8vo. Subscription $6 per year. (Foreign postage, 50 cents; Canadian 25 cents.)

Bulletin of the Johns Hopkins Hospital. Edited by FREDERICK B. BANG. Monthly. 8vo. Subscription $8 per year. (Foreign postage, 50 cents; Canadian, 25 cents.)

ELH. A Journal of English Literary History. Edited by E. T. NORRIS (Managing Editor) and D. C. ALLEN, G. E. BENTLEY, C. P. LYONS, ROY H. PEARCE, F. MICHAEL KROUSE and E. R. WASSERMAN. Quarterly 8vo. Volume 20 in progress. $4.00 per volume. (Foreign postage, 40 cents; Canadian, 20 cents.)

Hesperia. Edited by WILLIAM KURRELMEYER and KEMP MALONE. 8vo. Thirty-six numbers have appeared.

Human Biology: a record of research. The Editor is Dr. GABRIEL LASKER in association with Dr. JOSEF BROŽEK, Dr. BENTLEY GLASS, Dr. DONALD MAINLAND, Dr. JAMES N. SPUHLER, Dr. WILLIAM L. STRAUS, JR. Quarterly. 8vo. Volume 25 in progress. $5 per volume. (Foreign postage, 35 cents; Canadian, 15 cents.)

Johns Hopkins Studies in Romance Literatures and Languages. H. C. LANCASTER, Editor. 8vo. Seventy-six numbers have been published.

Johns Hopkins University Circular, including the President's Report and Catalogue of the School of Medicine. Eight times yearly. 8vo. Gratis.

Johns Hopkins University Studies in Archaeology. DAVID M. ROBINSON, Editor. 8vo. Thirty-nine volumes have appeared.

Johns Hopkins University Studies in Education. Formerly edited by FLORENCE E. BAMBERGER. 8vo. Thirty-five numbers have appeared.

Johns Hopkins University Studies in Geology. JOSEPH T. SINGEWALD, JR., Editor. 8vo. Sixteen numbers have been published.

Johns Hopkins University Studies in Historical and Political Science. Under the direction of the Departments of History, Political Economy and Political Science. 8vo. Volume LXX in progress.

Modern Language Notes. Edited by H. CARRINGTON LANCASTER, WILLIAM KURRELMEYER, KEMP MALONE, CHARLES R. ANDERSON, ARNO SCHIROKAUER, EARL R. WASSERMAN, and DON CAMERON ALLEN. Eight times yearly. 8vo. Volume 68 in progress. $6 per volume. (Foreign postage, 50 cents; Canadian, 25 cents.)

Reprint of Economic Tracts. Founded by J. H. HOLLANDER. Five series were issued.

A complete list of publications will be sent on request

Completely Revised And Enlarged

NUMERICAL MATHEMATICAL ANALYSIS

2nd Edition (1950)

By James B. Scarborough

Designed to serve as a textbook and reference work for students and workers in mathematics, statistics, engineering, and the exact natural sciences, Dr. Scarborough's revised and enlarged second edition sets forth logically and clearly the most important principles and processes used for obtaining numerical results. Means for estimating their accuracy are propounded, sufficient theory is given to show their soundness, and the limitations and pitfalls of these methods are pointed out.....$6.00

TABLE OF CONTENTS

The Johns Hopkins Press Baltimore 18, Maryland